The role of working memory in following instructions

Tianxiao Yang

PhD

University of York

Psychology

October 2011

Abstract

How do we follow instructions? Research has suggested that working memory may play an important role. This thesis explored the involvement of working memory in following instructions using dual tasks known to selectively disrupt the operation of the visuospatial sketchpad, phonological loop, and central executive components of the Baddeley and Hitch (1974) model of working memory. Across a series of seven experiments, working memory was found highly involved in encoding instructions. On the basis of these findings it is concluded that the central executive involvement was found to be most substantial, supporting the encoding and maintenance of sequences of actions. The phonological loop appears to play a general supporting role in registry and maintenance of verbal instructions. The contribution of the visuospatial sketchpad appears to be to encode and bind visual and spatial cues in an action, as well as retaining the sequence of actions, possibly via forming a map of locations of to-be-enacted objects. These roles of working memory were similar in following spoken and written instructions.

The secondary aim of the thesis was to investigate the action advantage in following instructions, which refers to the superior performance in enacting instruction sequences than simply recalling them verbally. This action advantage was established in both spoken and written instructions in a task paradigm containing rich visual spatial and motor cues, although was absent in a computer-based task involved limited actions upon abstract shapes. As the action advantage was not selectively impaired by the concurrent tasks employed in these experiments, its origins are unlikely to be in working memory. It is therefore concluded that working memory contributes substantial to the following of instructions, but it is not the source of the action advantage present in a rich task environment.

Contents

List of tables

List of figures

Acknowledgements

I am extremely grateful to my supervisor Susan Gathercole for her guidance and support during the course of my research. Her influence on my work has been more than merely supervising my research; she also sets a shining example of efficiency and excellence in research, leadership and life. I am also grateful to Alan Baddeley and Graham Hitch, who keep inspiring me in various ways, and I would also like to express my gratitude to Richard Allen for his insightful suggestions and advice about my research, especially regarding the design of experiments. Moreover, I would like to thank Silke Goebel, the third member of my research committee, for her encouragement and support, especially whilst I was writing up my thesis. Finally, I would like to thank Robert Logie and Tom Hartley, my examiners, who provided some excellent suggestions for the improvement of my thesis.

I also owe a great debt of gratitude to the Department of Psychology and the University of York. Not only did they support me financially by awarding me a department scholarship and an overseas studentship, they have provided an academic atmosphere which has instilled in me an enthusiasm for continuing research. I feel truly fortunate to have spent the last three years here. Additional thanks go to our excellent and supportive working memory group members, and especially to Shinmin Wang, for numerous informative and stimulating discussions about both research and life.

Last but not least, I thank my parents and Xiaodong Zhang for their dedicated love and everlasting patience during my three years' journey. This thesis is affectionately dedicated to them.

Declaration

I hereby declare this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, nor material which has been accepted for the award of any other degree or diploma at the University of York or any other educational institution. I also declare that the intellectual content of this thesis is the product of my own work and that it contains nothing that is the outcome of collaborative work.

Chapter 1

Introduction

A mind without instruction can no more bear fruit than can a field, however, fertile, without cultivation.

– Cicero

An overview

Knowledge exists to be imparted and one way of achieving this is by giving direct and specific instructions. From the perspective of learners, performing actions to command is also a human capacity that plays a key role in supporting everyday activities, e.g., cooking new dishes by following a complex recipe, remembering an instructor's commands when learning to drive, installing software on a computer after reading instructions on a website, and following teachers' commands in a classroom. These scenarios share one commonality: that is, all require remembering a series of action steps in sequence and performing them across a short-term period of time.

Indeed, following instructions is a complex cognitive process that involves multiple cognitive functions, such as perception, comprehension, memory and action. Because instructions typically guide actions that take place across time, the individual must remember the detailed content of the instruction at the same time as monitoring ongoing performance. This ability is associated with working memory (Engle, Carullo, & Collins, 1991; Gathercole, Durling, Evans, Jeffcock, & Stone, 2008), a limited capacity system that enables us to hold information in mind and manipulate it as necessary for a brief period of time. In particular, the flexible nature of working memory, and its vital role in learning and dealing with complex situations are both compatible with the demanding learning scenarios in which instructions are commonly received and followed.

Surprisingly little is known about either the cognitive processes that allow us to follow such commands or the ways in which instructions can be most effectively transmitted. This thesis aims to explore how working memory functions to support the ability to follow instructions. An understanding of the role of working memory in following instructions may provide insights into the difficulties people have in this process; for example children with low working memory span find it difficult to keep up with teachers' commands (Gathercole & Alloway, 2008). Besides the implications for education, understanding the cognitive process and the limits of people's ability to follow instructions may also help optimize the design of such instructions, and facilitate their efficiency and effectiveness in conveying messages.

The review of literature will start by introducing previous research on instructions presented in different forms of media, including written, spoken and demonstration as often reflected as imitation, followed by theories summarizing the characteristics of learning and the principles of teaching in a multimedia environment. Because actions are the basic elements that make up a typical instruction, the cognitive process, including planning and execution of actions, will be reviewed, along with two effects, namely the action advantage and the subject-performed task effect. Working memory is the focus of the current study; therefore working memory models will be presented, with a detailed description of the multi-component working memory model, which was used as the framework of this study. Given that following instructions is a complex task involving multiple cognitive functions, the interactions between perception, language, memory and action will also be covered briefly. This is followed by the presentation of some pioneering studies on the relationship between working memory and the ability to follow instructions. Finally, a summary of these findings and the structure of the thesis will be presented at the end.

Following instructions

The pioneering exploration of comprehending and using of instructions in everyday life was carried out by Wright (1978). Wright noticed that housewives paid little attention to the manuals of home appliances but preferred either fiddling with the appliance by themselves or watching it demonstrated by another. Another observation was that a question-and-answer dialogue seemed to suit people better than step-by-step procedural instructions. However, in the above situations, people might already have some background knowledge. In contrast, in situations of learning a new sequence of actions, step-by-step instructions may have the advantage of guiding people’s actions correctly and smoothly. Based on these findings, Wright (1978) proposed a three-tier control system for designing effective instructions. This starts by taking the reader’s needs and perspectives into account, and only after knowing the content that concerned readers most can the designers begin to think of the best way of communicating information, and the final step is to evaluate its effectiveness.

Written instructions

Conveying information through written text is efficient because massive amounts of detailed information can be encompassed in only a few lines or pages. Moreover, written instructions are less constrained by time and space, and can be transmitted rapidly especially in a time when electronic media is widespread. For example, it is common to see step-by-step text manuals for troubleshooting on the internet. Early research into comprehension noted that syntax affected the speed of comprehension and verification of sentences (Seymour, 1974). In a series of experiments, participants read instructions like ‘draw a circle above the square’, and executed the commands by drawing on a paper (Wright & Wilcox, 1978). It was found that participants focused on distinct psychological processes at different times. Participants assigned the surface structure segmentation during reading, assigned the locative features stated by propositions and planned the order of output thinking period, and were monitoring the output of drawing during the recalling stage. Another finding of this study was that sentences using main clauses like ‘draw A above B’ required less time in reading and thinking than sentences with an implicit embedded clause, such as ‘above B draw A’; this may be because that the embedded sentences were less likely to be segmented than the main-clause sentences. Nevertheless, the two types of sentences showed no difference in drawing time.

Interestingly, Wright and Wilcox (1978) also noticed that people tended to carry out the actions in the same sequence as the items mentioned. For example, when reading ‘draw a circle with a square above it’, participants drew the circle first. This is consistent with an earlier finding that people remembered the sentence better when the order of mention was same as the order of the event (Clark & Clark, 1968, cited in Wright 1978). This led Wright (1978) to suggest writing instructions according to the sequence of actions; for example, ‘do A before doing B’ is a superior form to ‘before doing A, do B’. It should be noted that the above research focused on the process of comprehending the content of instructions when the syntax was varied rather than how people memorized these instructions.

Spoken instructions

Compared to written instructions, spoken instructions are more flexible and convenient to produce and are common ways of giving commands in a face-to-face scenario, such as giving oral guidance in a classroom or tutoring athletes to improve their motor skills. Moreover, research has indicated that speech can be automatically registered into the phonological loop, a cognitive function that stores phonological information (Baddeley, Lewis, & Vallar, 1984; Hanley & Broadbent, 1987); therefore listening to speech should require less effort than reading. Another advantage of spoken instructions is their ability to work simultaneously with a visual system to guide actions (Henderson & Ferreira, 2004). In contrast, when people are following written instructions, they cannot both read and follow objects in space using the same visual system at the same time.

Nonetheless, spoken instructions can have drawbacks. A major one is that instruction receivers have less control over the speed of spoken commands, whereas people can read written instructions at their own pace. The ability to hold and process information is an ability known as working memory. In a widely-accepted working memory model by Baddeley and Hitch (1974), the phonological loop is one of the components which contains both the phonological store and the rehearsal mechanism (for more details see later section on the phonological loop). One characteristic of the phonological store is its rapid decay, and because items are usually chained in such a way that an item primes the next item (Ebbinghaus, cited in Baddeley, 2007), one step loss can sometimes lead to the loss of all subsequent steps. To prevent this catastrophic loss, people tend to rehearse the instructions. Rehearsal is considered to be a relatively automatic and economic way of doing this because speech code itself involves motor aspect, and there was assumed to be a direct mapping between speech perception and speech production (Liberman & Mattingly, 1985; Wilson, Saygin, Sereno, & Iacoboni, 2004). Therefore, repeating the instructions to oneself is a natural and convenient way of retaining spoken instructions.

Imitation

Before mastering language, imitation is an important way of learning, as both newborns and animals have shown the ability to imitate (Meltzoff & Moore, 1983, 1989; Whiten, 1998). Compared to following spoken and written instructions, imitation is a relatively automatic behaviour; for example, people often imitate each others' behaviour unconsciously during conversation (Chartand & Barge, 1999).

Direct mapping

The early tendency of imitating in infants and animals implies an automatic and direct mapping between the observed action and one's own action. This direct mapping has been validated by the discovery of mirror neurons in monkeys (Carey, 1996; Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996) and corresponding brain areas in humans (Buccino, et al., 2001; Decety, 2002; Grèzes & Decety, 2001; Grafton, Fadiga, Arbib, & Rizzolatti, 1997; Iacoboni, et al., 1999; Strafella & Paus, 2000). The mapping is considered as 'resonance' in the motor areas as soon as the visual input of the observed motions is presented no matter whether an action is executed or not, and no matter whether the action is meaningful or meaningless. It is hence inferred that the purpose of the resonance is to generate a representation encoding the other's action for future reproduction of the observed behaviour (Rizzolatti, Fadiga, Fogassi, & Gallese, 2002).

The direct mapping hypothesis was corroborated by the physiological evidence in humans that when people were observing the actions of others, motor-evoked potential was increased, leading to the same muscle activity as when they were executing those observed actions themselves (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995). There was also evidence from behavioural experiments. For instance, eye movement patterns during the perception of others' actions and the execution of one's own actions was proved to be similar. Moreover, the gaze behaviour was found to be proactive rather than reactive, suggesting the action planning during observation was automatic (Flanagan, 2003). The automatic planning was also corroborated by the proactive ideomotor movement of correction when participants were viewing a missing-target rolling ball (Prinz, 2002). Importantly, Prinz (1997) emphasized that this common coding for perceived and planned actions occurs at cognitive levels rather than a perceptual or an action level. This was further supported by the evidence of a common neural network shared during the observation, simulation, and execution of an action in a recent meta-analysis (Grèzes & Decety, 2001).

Active intermodal mapping mechanism

However, the direct mapping account cannot explain the deferred imitation observed in infants, which implies that an action representation is formed and maintained during the delay (Meltzoff, 1988; Meltzoff & Moore, 2002). Based on research into the imitative behaviours of infants, Meltzoff and Moore (2002; 1997) proposed the active intermodal mapping mechanism, in which the representation is a supramodal act space that unites inputs from different sensory modalities during observation. According to the common coding of perception and action in the direct mapping theory, the representation is supposed to be formed instantly without need of learning. When seeing the visually specified goals, infants are primed to attempt the imitative actions and receive proprioceptive feedback from their own actions, which are then matched with the supramodal representation of actions formed during observation. There is evidence suggesting that this representation is not a fixed one-to-one copying, but a context-free one that can be used flexibly in new settings (Barnat, Klein, & Meltzoff, 1996; Hanna & Meltzoff, 1993; Klein & Meltzoff, 1999).

Working memory in imitation

Although the active intermodal mapping mechanism suggests an intermediate representation interpreting the maintenance during delay, it does not specify the memory process involved. Rumiati and Tessari (2002) found that a concurrent motor suppression task interfered with imitating actions like ‘to comb, to brush, to eat’, and hence suggested the involvement of a motor storage system in maintaining actions. In contrast, there was no interference effect from a concurrent spatial tapping task, suggesting that there is little contribution from spatial storage in imitating movements. Moreover, it has been found that occupying working memory facilitates a primitive predisposition to imitate, implying the involvement of executive resources for inhibition (van Leeuwen, Baaren, Martin, Dijksterhuisa, & Bekkering, 2009). This inhibition may be particularly important when the imitative action has to be delayed. In addition, neuropsychological evidence has suggested that working memory is used during mental simulation; that is, working memory facilitates the construction of a dynamic motor representation by retrieving spatial and kinaesthetic information as well as serial plans of actions from long-term memory (Decety & Ingvar, 1990).

Dual routes

One hypothesized benefit of mental simulation is that once the motor representation becomes familiar, retrieval of actions from long-term memory become faster, implying there is another route besides the temporary and active storage using working memory. Thus, a dual-routes theory was proposed by Rumiati and Tessari (2002). They noticed that people were better at imitating meaningful actions than at mimicking meaningless actions. They explained the results by the different routes used during imitation. The semantic route is the long-term representation of familiar actions whereas the direct route uses a visuomotor conversion mechanism bypassing the long-term memory. The meaningless actions can only use the direct route by analyzing and parsing the movements into chunks to be stored temporarily in the working memory system. In contrast, meaningful actions employ the semantic route by retrieving the actions as a whole from long-term semantic memory, hence preventing the overload of the working memory system (Tessari & Rumiati, 2004). The dual-routes theory was corroborated by the neuropsychological evidence (Rothi, Ochipa, & Heilman, 1991, cited in Goldenberg & Hermsdörfer, 2002; Tessari, Canessa, Ukmar, & Rumiati, 2007), as well as evidence from a neuroimaging study (Rumiati, et al., 2005) in which with meaningful actions activated mainly the ventral stream, which specializes in semantic processing, in contrast to the dorsal stream, which deals with visuospatial transformation for meaningless actions.

Imitation and goals

Imitation also serves as a medium for understanding the intention state underlying the task or context (Tomasello, Kruger, & Ratner, 1993). Infants show gradual modification of imperfect imitative actions in order to achieve a resemblance to the observed actions (Meltzoff & Moore, 1994, 1997), and toddlers understand adults' goals by acting out the intended action when observing an unsuccessful attempt (Meltzoff, 1995). Two-year-olds imitate more of causally related events than arbitrary events while ignoring the irrelevant steps, suggesting that they have already grasped the structure and hierarchy of the actions (Bauer & Mandler, 1989; Bauer & Shore, 1987). With increased experience of perceiving the actions of others and the expansion of one's own behaviour repertoire, adults become used to analyzing others' actions in terms of goals and forming hierarchical plans (Travis, 1997).

Based on the observation and analysis of animal behaviours, Byrne and Russon (1998) proposed the hierarchical organization of imitation. They emphasized that imitative learning is organized hierarchically, and imitation mainly occurs at program level, a hierarchical layout of a behavioural 'programmer' rather than at the action level containing detailed and linear specification of sequential acts. This goal-directed view of imitation has been supported by Bekkering et al (2000). According to their view, imitation entails representing an observed behaviour as a set of goals, which subsequently drive the construction of an action pattern. The goals are hierarchical, and multiple goals compete for limited processing capacity; therefore the higher hierarchical goals are reproduced at the expense of lower goals. For example, when imitating complex tasks involving multiple goals such as objects and multiple movement paths, both adults and children made more errors of movement than the errors related to the target objects. This is because their focus on the higher-level goals, like objects, was at the expense of omitting the lower-level targets, such as movements (Gattis, Bekkering, & Wohlschläger, 2002).

Nevertheless, there are times when imitation can occur without understanding the process and purpose of the imitative actions. In fact, these 'mindless' imitative behaviours serve to help gain a fuller understanding of others' motivations and intentions during the imitation process (Byrne, 2002), such as the ability to imitate in young children before they are able to understand the intentions of adults (Meltzoff, 1988, 1995).

Comparing different types of media

Early research took an interest in the factors that influence the process of following instructions, such as delivery media, presentation modes and the sensory modality (Fleming & Levie, 1993; Mayer, 1997). However, all the research that set out to test the effectiveness of conveying information by comparing different media failed to find consistent results or conclusions (Clark, 1983, 1994; Clark & Salomon, 1986; Mayer, 1989; Mayer & Anderson, 1991, 1992; Mayer & Gallini, 1990; Mayer & Sims, 1994; Salomon, 1979/1994; Wetzel, Radtke, & Stern, 1994). Mayer (1997) hence suggested that effective instructions may depend more on factors other than the media.

Cognitive theory of multimedia learning

Based on the research into learning meaningful materials in a multimedia environment, Mayer (1997) proposed the cognitive theory of multimedia learning (CTML). This theory considers learners as active knowledge constructors rather than passive information recipients. The learning process involves four steps. First, multimedia presentations, like words or pictures, enter the senses, i.e. through the eyes and ears. Second, these senses select the modality-specific information. Third, the selected information is mentally organized in coherent verbal and pictorial representations. Finally, integration occurs both between the two types of representations, and between these and the representations of existing knowledge (Clark & Mayer, 2008).

CTML assumes dual processing channels and a limited working memory capacity, and emphasizes that successful learning involves active processing and transfer. The model itself does not provide revolutionary perspectives on learning, and Mayer himself also admitted that this theory draws on several other theories, including Wittrock’s generative theory (Wittrock, 1974), Paivio’s dual coding theory (Paivio, 1986b), and also Baddeley and Hitch’s multicomponent working memory theory (Baddeley, 1992). Nonetheless, several effects discovered using this model have been influential in the field of education, especially in the newly emerged e-learning environment.

Effects in multimedia environment

The multimedia effect suggests that information from multiple media is beneficial. For example, including visual information in instructions improves learning, provided that narration is coordinated with animation and text is coordinated with illustrations (Fletcher & Tobis, 2005). This is because that encoding the material both visually and verbally requires the mental construction and integration of the pictorial and verbal representations. This active mapping of the two systems helps build a coherent mental representation that facilitates better learning (Clark & Mayer, 2008, p.65; Mayer, 1997).

One hidden premise of the multimedia effect is the multiple modality view of working memory, suggesting that the effective size of working memory can be increased when multiple modalities are used. Following this reasoning, multiple sources of information should be presented through different modalities to avoid any traffic within the same modality. Therefore, when explaining graphs or animation, it is better to accompany the visuals with speech than with on-screen text, which uses the same visual modality. This modality effect has been observed in many studies (Craig, Gholson, & Driscoll, 2002; Mayer, Dow, & Mayer, 2003; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Moreno, Mayer, Spires, & Lester, 2001; Mousavi, Low, & Sweller, 1995; O'Neil, et al., 2000). In addition, a recent meta-analysis (Ginns, 2005) indicated two moderators of the modality effect; one is the element interactivity, with a larger modality effect for materials that contain a high interactivity of elements. The other is the pacing of presentation, with a smaller modality effect when the pace of presentation can be controlled by learners rather than being controlled automatically by computer. These two moderators imply that the modality effect is more likely to occur in learning situations that require a high working memory load.

The contiguity effect argues that the multimedia effect is effective when information from different media is close in time and space, that is, both spatial proximity and temporal continuity facilitate learning (Clark & Mayer, 2008). Evidence has shown that improved learning occurs when corresponding graphics are placed near the printed words (Chandler & Sweller, 1991; Mayer, 1989; Mayer, Steinhoff, Bower, & Mars, 1995; Moreno & Mayer, 1999; Paas & van Merrienboer, 1994; Sweller & Chandler, 1994; Sweller, Chandler, Tierney, & Cooper, 1990) and when narrations and corresponding animations are presented simultaneously (Baggett, 1984; Baggett & Ehrenfeucht, 1983; Mayer & Anderson, 1991, 1992; Mayer, Moreno, Boire, & Vagge, 1999; Mayer & Sims, 1994). Understandably, it is easier to build connections between verbal and visual representations when both of them are still in the short-term memory. In contrast, the spatially or temporally separated materials require extraneous processing that is unrelated to the instructional goals, thus burdening the limited working memory, leaving less capacity for integrating goal-relevant materials, and hence impairing the learning (Clark & Mayer, 2008, p. 90; Sweller, et al., 1990).

The redundancy effect and coherence effect highlight the importance of avoiding the extraneous processing. For example, some designers like to include printed text with a narrated graphic to accommodate different learning styles. However, this may lead to split attention because learners might divert their attention to the printed words and therefore they pay less attention to the accompanying graphics. In addition, people tend to compare the printed text with narrations, and printed words and pictures will compete for visual processing resources. In short, redundant materials tend to cause extraneous processing therefore harming the learning (Craig, et al., 2002; Kalyuga, Chandler, & Sweller, 1999, 2000, 2004; Mayer, Heiser, & Lonn, 2001; Moreno & Mayer, 2002; Moreno & Mayer, 2000b).

Similarly, extraneous audios, graphics and text should be omitted, as extraneous processing irrelevant to the learning aim hampers study even if it is added for additional interest (Mayer, et al., 2001), for expanding on key ideas (Mayer, Bove, Bryman, Mars, & Tapangco, 1996), or for technical depth (Mayer & Jackson, 2005). This is because the arousal may divert attention away from the original content of the learning material that should be the focus of attention (Clark & Mayer, 2008), and extraneous details may prime the theme of irrelevant existing knowledge background, leading to inappropriate integration (Harp & Mayer, 1998). For instance, detailed colour drawings are found to be less effective than simple line drawings (Butcher, 2006; Parkhurst & Dwyer, 1983). Sometimes, even presenting sound that relevant to the learning task can impair retention and transference of the knowledge (Moreno & Mayer, 2000a).

Cognitive load theory

The cognitive load theory is another influential theory that relates cognitive functions and instructional design. The theory asserts that instruction information organization depends on individual working memory capacity when dealing with novel tasks, and when it exceeds the limits, the performance dropped; but for experts, long-term memory like schemas facilitate efficient organization (Jong, 2010; Sweller, 2004; Tindall-Ford & Sweller, 2006).Sweller (2004) emphasized the role of long-term memory, suggesting it determines the characteristics of both working memory and sensory memory, and that individual background knowledge and working memory capacity should therefore be considered when designing instructions.

As its name reveals, this theory emphasises the importance of cognitive load. There are three types of load. First, the intrinsic load is defined as the number of elements to be held and manipulated simultaneously in order for a particular process to be learned. This is determined by the complexity of the learning material and is assumed to be beyond the control of the instructors. By contrast, the extraneous load is the unnecessary cognitive demands imposed by instructional design; for example, presenting both chart and bar figures for the same data is redundant, and should therefore be decreased following the principles of cognitive load theory. The germane cognitive load is the load devoted to the mental organization of newly-learnt materials and the integration of these with existing knowledge, a result of active effort in organizing relevant materials (Mayer & Moreno, 2010; Moreno & Park, 2010). Unlike the other two, germane load has a positive relationship with learning, and is increasable by appropriate instructional design. For example, in the case of learning how to use a computer spreadsheet application, after having held the prerequisite schemas of the procedure, students who were engaged in imaging outperformed those who focused on understanding and remembering (Cooper, Tindall-Ford, Chandler, & Sweller, 2001).

These three loads – intrinsic, extraneous and germane – add together to make up the total cognitive load, which consumes part of the working memory resources; what is left is the free capacity. The aim of the cognitive load theory is to optimize the free capacity through decreasing the extraneous load and to facilitate learning by increasing the germane load.

Effects and techniques

Effects in the cognitive load theory that relate to the extraneous load overlap with some effects in the cognitive theory of multimedia learning (Mayer, 2005); I thus focus here on the effects and techniques that aim to increase the germane load.

The imagination effect occurs as a result of improved learning when learners imagine or mentally practice a procedure or concept being learned, compared with simply studying the material (Cooper, et al., 2001; Ginns, Chandler, & Sweller, 2003; Leahy & Sweller, 2004; Leahy & Sweller, 2005; Tindall-Ford & Sweller, 2006). This effect occurs only when learners have sufficient prior knowledge; if this is not the case, imagining in this way leads to poorer learning than studying, such as in the initial stage of learning for novices (Kalyuga, Ayres, Chandler, & Sweller, 2003). This is because the schemas from the long-term memory make the mental manipulation of elements feasible and easier. Moreover, imagining the task helps learners focus on the crucial entities and eliminates the unnecessary searching and checking process during learning (Tindall-Ford & Sweller, 2006).

As stated before, the numbers of elements to be processed simultaneously in the working memory is the primary source of the intrinsic load, which depends on the nature of the material and determines the task difficulty. The load of element interactivity is also contingent on the schemas held in the learners' long-term memory, and differs for novices and experts. During the learning process, the lower-order schemas become an element of a high-order schema that can act as a single element (Kalyuga, 2010). Therefore, in situations when people are learning unfamiliar materials with a high level of element interactivity, learning can be enhanced by segmenting the material by first teaching isolated elements then introducing an interactive version, rather than introducing interactive elements in the beginning (Kester, Paas, & van Merrienboer, 2010; Pollock, Chandler, & Sweller, 2002).

Individual difference and limitations

Most effects in the cognitive load theory show large individual differences, reflecting mainly in the domain of information processing such as intelligence and prior knowledge. For example, learners with high spatial ability benefit more from the temporal contiguity effect (Mayer & Gallini, 1990; Mayer & Sims, 1994; Mayer, et al., 1995). A person with larger schemas storage tends to be efficient in organizing information, and the instructional methods that are effective for novices may become less effective for experts, including the contiguity effect (Mayer & Sims, 1994), the redundancy effect (Kalyuga, Chandler, & Sweller, 1998), and the multimedia effect (Kalyuga, et al., 2000; Kalyuga, Chandler, & Sweller, 2001; Kalyuga, Chandler, Tuovinen, & Sweller, 2001; Tuovinen & Sweller, 1999).

Moreover, it has been suggested that learners who have more prior knowledge also tend to apply deeper and more effective strategies, and use working memory resources more wisely than novices (Plass, Kalyuga, & Leutner, 2010). Learners who are poor self-regulators have been found to learn better in a program-controlled condition than in a learner-controlled condition, whereas the higher self-regulators showed no difference in both conditions (Eom & Reister, 2000). These results suggest more guidance is needed for low-regulators.

The main purpose of the cognitive load theory is to optimize the way of presenting novel information to accommodate the limited working memory capacity, using a range of principles in order to reduce unnecessary working memory load and facilitate change in the long-term memory (Sweller, 2010). However, the principles and effects of the cognitive load theory came from limited learning scenarios, mainly scientific material; extending these to other domains is therefore questionable, and the long-term consolidation of these effects has not been tested (Burnken, Plass, & Moreno, 2010). Moreover, the cognitive load theory does not explain how information is processed and represented (Horz & Schnotz, 2010). It is therefore worth looking back to an earlier theory – the dual coding theory.

Dual coding theory

In order to understand the way in which people represent environmental stimuli, Paivio (1971) proposed the dual coding theory which originated in the view that distinctive experience gives rise to specific characteristics in different domains in order to serve corresponding functional or adaptive goals. There are two independent but also cooperative subsystems: the verbal system and the nonverbal system. The verbal system specializes in processing information related to language with internal representational units called logogens, in contrast to the nonverbal or imagery system, which specializes in dealing with nonlinguistic information with units called imagens (Paivio, 1990).

The dual coding theory explains the between-system relations as referential connections, where one system can trigger the other system given the underlying structural connection. According to Paivio (1990), the between-system activation is not automatic but conditional, depending on the interaction between stimulus context and the functional strength of the referential interconnections, which are determined by previously activated representations. Evidence has shown that the referential connection process of a verbal-to-image representation and an image-to-verbal representation is asymmetric. Representing an image verbally elicits a range of names for this image which vary in their probability, whereas a word stimulates a prototypical image (Snodgrass & Vanderwart, 1980).

The verbal and imagery systems differ in several ways. From a structural point of view, the information units in verbal systems are connected associatively and logically hierarchical, corresponding to the linguistic hierarchy based on natural categories; the nonverbal system reflects the world in continuity and allows nested imagery to expand into a broader setting. Functionally, the verbal system is constrained by sequential processing whereas the nonverbal system contains simultaneously available information that can be accessed in various perspectives, which means that it is not sequentially constrained by the representational structure. Moreover, Paivio (1990) suggested that an accompanying motor process is involved in the nonverbal transformation.

A naturally following question from this is what determines the type of representation to be formed upon instructions. Paivio (1990) suggested that it is a joint function of stimulus situation and individual differences. Specifically, empirical studies found that imagery is likely to be evoked by concreteness (e.g. objects and pictures) and instructions relating to an image; conversely, verbal representations are likely to be activated when using words with high verbal associations as stimuli and the tasks which demand verbal processing, in particular when instructions are given to carry out a task verbally. Nevertheless, Paivio admitted that activation can involve one or both, or even combine both types of representations, which seems to have additive effects on recall. However, given the various influencing factors mentioned above, it is hard to predict the exact representation. Besides the characteristics of stimuli, individual factors, such as a preference for verbal or imagery and cognitive ability, can also influence the probability of representation activation.

The dual coding theory emphasizes the benefits of concreteness and imagery, as memory performance increases from abstract words to concrete words and to objects (or pictures). This is because presenting an object is likely to trigger a covert naming process, resulting in a dual verbal and nonverbal memory trace, whereas the abstract words are difficult to imagine and therefore less likely to be dually coded (Paivio, 2006). Compared to the sequential processing of verbal representations, the synchronous and integrative properties of imaginal representations tend to facilitate associative learning, and the imaginal codes are more beneficial in terms of mnemonic value, with an estimated ratio of 2:1 compared with verbal codes. Nevertheless, a verbal mechanism has its own strength in controlling the mental process of manipulating images, and this control can sometimes be carried out without awareness (1990).

Summary

Instructions can be conveyed by different types of media, and presentation modes, and each has its own pros and cons. Several theories have been developed to decrease the cognitive load. Some effects and principles beneficial for learning in a multimedia environment were identified.

It should be noted that these principles were developed from the scenarios of studying meaningful material measured by the effectiveness of knowledge transfer, which might be different from learning situations that require retention or operations. For example, temporal contiguity effect obtained in the measure of problem solving failed to occur in a retention test (Mayer & Anderson, 1992). Therefore, it seems that different learning purposes lead to different cognitive processes (Aaronson & Scarborough, 1976). Most theories reviewed above are concerned mainly with the encoding stage; the next section will thus focus on the output aspect, concentrating on one of the most common forms of output, the actions.

Actions

Norman and Shallice distinguished actions that are relatively automatic, like habits and schemas, and those less automatic actions that demand control by the supervisory attentional system (Norman & Shallice, 1983, 1986). Different from the automatic routines, which, once being started, unfold without wilful control, the more voluntary types are often newly-learnt actions which require additional steps like action planning.

Indeed, action is a complicated, dynamic, and a competitive process that links the sensory information with the intention to move. Recent neuroimaging research has provided evidence for this intricate nature of action, reflecting in the graded activation and interactions of cortical and subcortical brain areas (Purves, et al., 2008). Specifically, the premotor and posterior parietal cortex is modulated by motivation, and the supplementary motor area programs the action sequence that is later issued by the primary motor cortex as motor program commands. These commands go through the basal ganglia, which interacts with the motor cortex and plays a gating role by inhibiting potential movements until the appropriate circumstance occurs. The parietal cortex then integrates the visual and motor positions, and produces coordinated movements of the eyes and hands. During the execution of actions, the cerebellum is employed to help coordinate movements and correct errors.

Action planning

One noteworthy characteristic of an action representation is its hierarchical organization. The goal of an action motivates and starts the action planning, which is the highest level of motor representation. Action planning requires associating cues and movements, selecting motor schemas and organizing the temporal framework (Jeannerod, 1997a). Based on the distinction between routine and nonroutine activities, Stuss and his colleagues proposed a model to explain the action planning mechanism (Stuss, Shallice, Alexander, & Picton, 1995a). There are four components in this model, namely, the cognitive modules, schemata, contention scheduling, and supervisory system. The cognitive module contains basic operations, and is controlled by schemata. The schemata are routine programs of overlearned skills and the crux of the whole system. A schema receives activation from perceptual input, other schemata and also the supervisory system, and it produces output to the effector system and other schemata as well as providing feedback to the supervisory system. The schemata are organized hierarchically into a more complex routine, and sometimes need contention scheduling to control the competition between schemata.

In contrast to the three aforementioned components that relate to routine activities, the supervisory system manages the nonroutine activities, mainly through top-down activation and the inhibition of schemata, and it also adjusts the contention scheduling and monitors the schema activity. The supervisory system is especially useful in helping establishing new schedules in newly-performed actions, including selecting and activating a number of stored schemas, and organizing their modalities and time of expression so as to reassemble these schemata into a coherent action plan (Jeannerod, 1997a). There is also evidence suggesting this supervisory system in action planning is associated with the prefrontal cortex during the action planning (Jeannerod, 1997a; Purves, et al., 2008; Stuss, et al., 1995a). Another way of having conscious control over newly-learnt actions is through verbal conceptualization; for example, it is observed that people often use inner speech to rehearse oral commands in the early stage of learning a new sequence of actions (Adams, 1971; Decety & Ingvar, 1990; Schmidt, 1975).

Mental practice

A more explicit way of action planning is mental practice, defined as the mental rehearsal of a task in the absence of simultaneous sensory input and overt physical movement (Driskell, Copper, & Moran, 1994). Here I focus on two types of imagery that are relevant to carrying out future actions, visual imagery and motor imagery. Visual imagery refers to the internal simulation of visual process, such as visualizing somebody else performing the actions (Engelkamp, 2001). Motor imagery concerns the internal simulation of motor process, such as imagining oneself performing the actions in the task environment. Motor imagery is quite common in preparation for actions and has been proved similar to the actual action in many ways, such as the similar durations between the mental simulation and the actual execution of an action (Decety, Jeannerod, & Prablanc, 1989; Parsons, 1994).

Mental practice has been found to boost performances of both cognitive tasks and physical tasks (Driskell, et al., 1994), and several accounts have been put forward to explain this benefit. Two major accounts are the inflow processing and the outflow processing. Based on the finding of accompanying electromyographic activity during a simulated motor act (Jacobson, 1932, 1973; Freeman, 1931, cited in Decety & Ingvar, 1990), inflow processing supposes a closed loop system requiring proprioceptive and peripheral feedback. It argues that mental practice causes minute innervations in the muscles, resulting in the kinaesthetic feedback, hence strengthening the motor program. However, later studies failed to replicate these innervations in the muscles during mental simulation (Driskell, et al., 1994; Jeannerod, 1997c).

The outflow explanation supposes an open loop system depending on a pre-planned serial movement sequence (Lashley, 1951, cited in Decety & Ingvar, 1990). It posits that the effect of mental practice happens at a higher cognitive and symbolic level, the programming and planning level, rather than at a lower perceptual or muscle level. This account predicts that mental practice is more effective in the early stage of motor learning, during which it contributes to the construction of cognitive plans (Schmidt, 1975). Simulation in the mind optimizes the mental plan and hence facilitates symbolic control over movement or learning. Moreover, mental practice was found to help refine the motor programming and control in the later stage of learning (Savoyant, 1988, cited in Decety & Ingvar, 1990). The outflow explanation has been supported by the bilateral transfer effect of a learnt motor response (Kohl & Roenker, 1980, 1983).

According to Decety and Ingvar (1990), mental simulation of actions as a cognitive modelling requires various cognitive components. When actions are experienced consciously, such as during a delay or whilst being disturbed, a construction of a dynamic representation is likely to be formed in the working memory, which combines spatial and kinaesthetic schemas in the long-term memory as well as the activation of serial plans of action. Indeed, prior knowledge or schemas were found to influence the effect of mental practice. Experienced learners benefited more than novices, and novices gained more on the cognitive than physical tasks, suggesting the importance of prior semantic knowledge for the effectiveness of mental practice (Driskell, et al., 1994).

Moreover, mental practice was similar to the actual execution of actions, as reflected in the large overlap of the activations in brain areas during action execution and simulation (Grèzes & Decety, 2001). The similar temporal organization in mental simulation and actions also suggests that mental practice might have helped set up timing for the actual performance (Decety, et al., 1989; Decety & Michel, 1989).

It is worth noticing that physical practice provided additional gains. Kohl and Roenker (1983) found larger unilateral than bilateral during physical practice, suggesting extra gains were made from physical practice. It is possible that actual practice provides proprioceptive and visual feedback of one's own action. The lacking of physical feedback can sometimes cause one to underestimate the difficulty met during execution; for example, the duration of imaged movements were shorter than actual movements when the required movements were difficult (Parsons, 1994).

Monitoring during action execution

Carrying out actions involves retrieval action plans, paying attention to external cues as well as monitoring one's own actions, and sometimes requires one to store intermediate goals. Importantly, matching the outcome of actions with goals means constant monitoring during execution. Jeannerod (1995, 1997d) therefore argues for a more dynamic monitoring process in contrast to a rigid test-operate-test-exit monitoring method (Miller, Galanter, & Pribram, 1960).

In Jeannerod's model (1995), an action is organized hierarchically, including intention, planning, programming and execution, ranked from the highest level to the lowest. Importantly, there is a control mechanism parallel with the main stream of the levels. At each level, operations performed are stored as motor memories, which are used as a comparator to compare the desired actions with the current state. The current state is not merely the visible results of the intended action, nor is it the simple sensory feedback, but a result of careful calculation. This is due to the need for minimizing correction delays of intended actions when unpredicted perturbations occur in the environment. Therefore, motor commands often look ahead in time by producing a forward model that estimates the outcome of actions without receiving feedback from the actual performance. This is also known as the corollary discharge in neurophysiology, which postulates that signals generated by the motor centres provide information about future movements before they are reaching the effectors.

At the same time, a model of sensory output was generated by a comparison of predicted and actual sensory feedback, and any difference regarded as sensory error was used to correct the state estimated from the forward model (Wolpert, Ghahramani, & Jordan, 1995). Sometimes, the alternation provided by the error feedback is insufficient to obtain the desired effect; thus, the program level remains activated with error signals propagating to a higher level, leading to the change of action plan or the setting up of a new plan. Only when the intended actions are completed are corresponding memories of the intended goals erased.

Importantly, the dynamic monitoring model emphasizes the intertwined stages of actions. Moreover, it argues that the internal model of actions has to continuously interact with the external world in order to provide speedy feedback for actions to be smoothly and correctly executed. This dynamic monitoring process is found to be assisted by the subvocalization, which serves as a means for maintaining strategic control of actions (Baddeley, 2003a; Baddeley, Chincotta, & Adlam, 2001).

Action advantage

A consistent finding in the following of instructions is the superior performance of recall by actually carrying out the actions than orally repeating the instructions. This action advantage exists in both children and adults, and in spoken as well as written instructions (spoken instructions, Allen, 2009; Gathercole, et al., 2008; written instructions, Koriat, Ben-Zur, & Nussbaum, 1990).

Koriat et al. (1990) were the first to notice the benefit of recalling a series of actions by performance rather than by oral repetition in a series of experiments. In their experiments, participants read instructions containing three or four actions common in everyday life, such as ‘lift the ashtray, move the stone, tap the eraser’. Participants were better at performing the instructions than repeating them sequentially verbatim. Moreover, in a crucial experiment, Koriat showed that expecting to perform actions led to superior oral repetition than if the participants were expecting oral repetition but were then required to perform the actions. This result was explained as representing the benefit of recalling through actions, rising from the encoding stage rather than the retrieval stage. Specifically, the representation underlying memory for future actions takes advantage of the imaginal-enactive properties of envisaged actions, which is superior to a verbal-based representation of abstract propositions that need to be translated into actions during retrieval. The superiority of an action representation to a verbal representation was based on the aforementioned dual coding theory (Paivio, 1990), in which the verbal system is constrained by sequential processing, whereas the nonverbal system allows access to information from various perspectives simultaneously. The reason why different representations were formed for different types of recall might well be the need to save the transformational cost between representations and output modalities. In a third experiment, the findings of action advantage and benefits in the encoding stage were extended to the long-term memory. Nevertheless, one difference between the two types of recall was observed in the output stage, which was that participants were more likely to repeat a previously communicated action in the oral repetition than in the action performance, implying that output monitoring is more effective for motor enactment than verbatim recall.

This action advantage was later replicated in an experiment requiring participants to remember a series of oral commands of actions (Allen, 2009). Allen agrees with Koriat et al. (1990) that the benefit of action recall lies in the encoding stage, but he argues that planning for actions facilitates the formation of an integrated multimodal representation involving phonological, visual and motor codes, whereas this multimodal representation is not present when an oral repetition is expected. The multimodal representation integrates elements from various channels into a coherent representation.

More importantly, the formation of a multimodal representation is an efficient way of connecting action intentions with the external world. This corresponds to Glenberg's idea that the mission of memory is to encode patterns of possible physical interaction with a three-dimensional world through conceptualization, which are constrained by our bodies (Glenberg, 1997). In addition, the utility role of objects and the visual features of objects are known to prime motor activity (Wilson, 2002), which also relates to the ecological and evolutionary perspective of the affordance of perceived objects (Gibson, 1977, 1986). These views are related to the notion of the importance of embodiment in serving memory.

Other factors may also contribute to the action advantage, such as motivation, experience, and feedback. For instance, actions usually relate more directly to the goals and produce visible outcomes; hence planning for actions is likely to evoke more active processing compared to simply repeating the commands in words, which have indirect or little effect on the external world. Importantly, these visual outcomes of one's own actions, along with the proprioceptive feedback, provide effective guidance for building representations for future similar actions. Moreover, our experiences of interacting with objects by using our hands start earlier than our abilities to describe them verbally; for instance, infants start learning by imitating actions in a matching-to-target process (Meltzoff & Moore, 1977; Meltzoff & Moore, 1994). All these possibilities indicate a closer and more over-practiced stimuli-response link for actions than for oral repetition.

All in all, these explanations and conjectures point to a superior representation for actions than for oral repetitions. It appears that a number of factors might influence the representation of instructions, and the construction of a representation is likely to vary with situations and is probably more complicated than we might expect.

The subject-performed task effect

Another action-related phenomenon is the subject-performed task (SPT) effect or enactment effect. It is the advantage of encoding actions by performing them during encoding. In a typical task, participants were presented with a list of mini tasks of simple actions, such as 'open the book', 'lift the hat' etc. One group of participants simply listened to the list of actions (the verbal task) whereas the other group listened while also performing these actions (the SPT task). A consistent finding was that performing the actions during encoding led to better free recall than only listening to the action phrases, i.e. the SPT effect. Moreover, the memory performance in the SPT conditions was better than the condition in which participants imagined carrying out these actions during encoding. The SPT effect was also larger than the benefit gained from observing an experimenter performing the actions during encoding (Engelkamp, 2001).

The SPT effect is assumed to be non-strategic, arising from an effortless encoding process, which leads to the automatic creation of robust representation in memory (Cohen, 1981). There is still heated debate over the underlying mechanism of the SPT effect, which might be due to the multi-modal and contextually rich encoding (Bäckman, 1985; Bäckman & Nilsson, 1984), the benefit of encoding motor movement (Engelkamp & Zimmer, 1985, cited in Engelkamp, 2008), or self involvement and experiential registration (Kormi-Nouri, 1994, cited in Zimmer et al., 2001). In particular, this self-involvement of actions relates to the recent emerging area of embodiment, which also emphasizes the importance of body and self-representation in action, language, and social interaction (Klatzky, MacWhinney, & Behrmann, 2008).

Recently, two studies explored the SPT effect in modulating the action advantage in following instructions. In one study, participants listened to commands requiring series of actions upon laminated cards of shapes, and in the SPT condition participants were asked to enact the actions in addition. The SPT task boosted the participants' performance at recall (unpublished data, from personal communication with Allen). Moreover, it improved the performance of oral repetition more than that of action recall, implying that encoding actions helped later verbatim repetition of these actions. This SPT effect can either be attributed to the mental practice that reinforced the multimodal representation during encoding, the benefit of motor coding, or perhaps both. The SPT effect in following instructions was also extended to children (Wojcik, Allen, Brown, & Souchay, in press). In that study, children listened to a series of spoken instructions containing actions upon colourful stationery. It was found that the children's recall performance in the SPT condition was superior to the conditions in which they only listened to the instructions.

Summary

Both the action advantage and the subjected-performed task effect indicated some benefits underlying the representation of actions. On one hand, action can arise from internally generated intentions, which serves as the goal that musters various cognitive functions and guides them to work cooperatively until its completion. From this point of view, action is a top-down hierarchical process that involves the actualization of a series of subgoals. On the other hand, action can be a bottom-up process, triggered by objects, and the stages of actions are intertwined in a dynamic way in order to cope with the changes in our environment.

For example, seeing a cup of tea may trigger the thirsty feeling that transfers into a goal of picking up the cup and drinking the tea, initiating the formation of a plan for a sequence of actions to achieve the goal. The movement of picking up the cup is monitored by a supervisory system. In the case of a contingent event, such as the cup being too hot to hold, the original planned action has to be delayed while the intention and the action plan are maintained until the impediment in the environment disappears. Besides the need for visuo-motor coordination, recognizing the object and activating the pragmatic knowledge requires semantic processing; in this case, identifying that the cup contains drinkable tea can alleviate thirst. Newly-learnt actions are likely to involve additional processing, such as the way of shaping hand gestures to accommodate an unfamiliar object, memorizing a novel action sequence and inhibiting the tendency of performing a similar routine action which would be unsuitable for this new task. All these functions, planning, maintaining and monitoring tend to load heavily on the working memory, which we shall turn to now.

Working memory

Working memory is a cognitive function that maintains and manipulates information. Various models have outlined the structure of working memory as well as explained its functions. Here, three influential models are reviewed, with the focus on the multi-component working memory model.

Early models like Atkinson and Shiffrin's modal model (1968) proposed a serial processing of information with three consecutive components. Environmental information is first registered by sensory memory and then flows to the short-term memory, which holds and manipulates the information, either leading this to an immediate output or storing it into the long-term memory. This model was criticized for its simple linear processing mode in its description of how information transfers from short-term to long-term memory (Baddeley, 2007). The report of a neuropsychological patient with intact long-term memory but impaired short-term memory learning ability indicated that short-term memory may not be the only buffer through which information can enter into the long-term memory (Shallice & Warrington, 1970; Warrington & Shallice, 1969). Another objection is the statement that longer duration of maintenance leads to better long-term memory, which was proved wrong by the finding that longer rehearsal time did not improve the recall (Craik & Watkins, 1973), and that what matters is the level of processing (Craik & Lockhart, 1972).

The embedded-processes model of working memory

The embedded-processes model of working memory emphasizes the links between perception, attention and long-term memory (Cowan, 2005). Working memory information comes from the hierarchically organized faculties including long-term memory, the subset of long-term memory that is activated, and the subset of activated memory that is in the focus of attention. The focus of attention is limited in capacity, with a magic number of four items being hold at one time (Cowan, 2001). By contrast, activated memory is assumed to be unlimited in its capacity but subject to limitation of time; that is, it decays as time passes (Cowan, 1988). The control and regulation of working memory is via the control of the focus of attention, which is controlled jointly by a voluntary central executive system as well as an involuntary and automatic recruitment of attention. The influence of attention can be seen in the various stages of memorization; for example, attention can enhance the coding process as well as changing the nature of perceptual encoding. Maintenance of information is viewed as keeping items in the focus of attention to continue activating them in the memory. Retrieval is also seen as a process that helps enter the correct items into the focus of attention while racing against forgetting in the activated memory.

In essence, Cowan views the effective working memory as the vehicle for the retrieval of all information relevant for the completion of a particular task; therefore, various mechanisms, including memory activation, attentional and executive, as well as long-term memory, all work together to function as an effective working memory system (Cowan, 1988). For example, many tasks in life, such as following instructions, involve novel combinations of stimuli, which require concurrent activation of all relevant information. Meanwhile, the focus of attention helps prolong the action to allow them to be bound, and these new combinations are likely to be stored as a new long-term memory trace. These newly-formed long-term memory traces can also be used as virtual short-term memory for other similar tasks, which is similar to the concept of long-term working memory (Ericsson & Delaney, 1999; Ericsson & Kintsch, 1995). In general, this embedded-process model of working memory provides a broad view of the functions of working memory in complex activities as its ability to bring together various cognitive components mechanisms simultaneously (Cowan, 1999).

The multiple component working memory model

The multiple component working memory model (Baddeley & Hitch, 1974) is a useful model that provides an invaluable framework for guiding the empirical investigation of complex cognitive activities. This model emphasizes maintenance as well as manipulation of information, and has well-established paradigms useful for exploring the cognitive constructs in a complex task. The most recent model consists of four components (Baddeley, 2000, see Figure 1.1). The central executive is responsible for directing attention and coordinating information within the working memory and across the cognitive system more generally. They are supplemented by temporary stores for verbal and visuospatial material; these are termed the phonological loop and the visuospatial sketchpad respectively. Although this model acknowledges that working memory retrieves information from the stored long-term knowledge relevant to the current task, it also focuses on how working memory supports learning by the manipulation and recombination of new material to allow interpretation and then the encoding of this into long-term memory (Baddeley & Logie, 1999). Compared to Cowan's model (2005), this model considers working memory and long-term memory as more functionally separable systems.

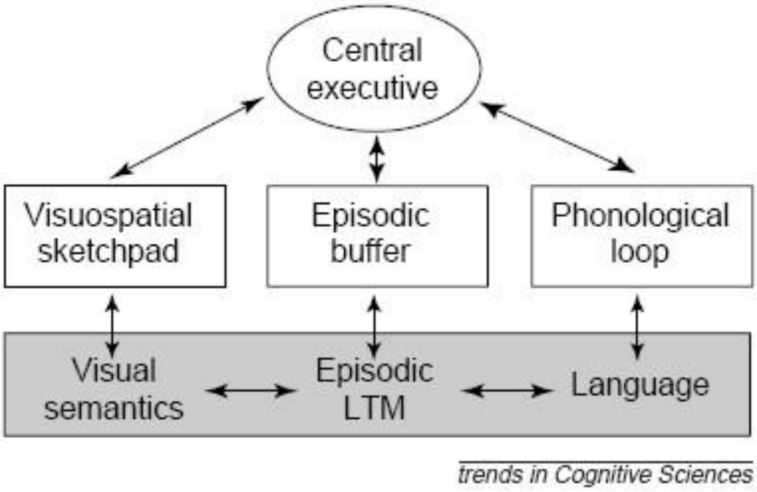

Figure 1.1 The multimodal working memory model (2000). The shaded area represents the long-term memory

The phonological loop

The phonological loop is a system specializing in storing verbal information, and is found to facilitate the early stage of word learning (Gathercole, 2006). It comprises a phonological store and an articulatory rehearsal process analogous to subvocal speech (Baddeley, 2003b). The separation of storage and rehearsal is supported by the neuropsychological evidence, reflected in patients with lesions affecting either storage or rehearsal (Vallar & Papagno, 2002). The separation is also reinforced by neuroimaging evidence that storage and rehearsal activate different brain regions. Specifically, storage activates the supramarginal gyrus in the left temporal lobes, whereas rehearsal activates the left frontal region (Broca's area) and the left premotor frontal regions (Henson, Burgess, & Frith, 2000; Paulesu, Frith, & Frackowiak, 1993).

The store maintains information in a phonological form and gives rise to the phenomenon known as the phonological similarity effect, in which the recall of lists of visually-presented items with distinct sounds such as W, X, K, R, Y is superior to the recall of a phonologically similar sequence such as V, B, G, T, C (Conrad, 1964; Conrad & Hull, 1964). The result also indicates that visual material containing verbal information can be transformed into a phonological store. In contrast, auditory sound gains access to the store automatically, proved by the interference effect of irrelevant speech (Hanley & Broadbent, 1987; Neath, Surprenant, & LeCompte, 1998), and the remaining phonemic similarity effect under articulatory suppression (Baddeley, et al., 1984; Longoni, Richardson, & Aiello, 1993). The capacity of the phonological store is limited and items held in the store eventually fade away. The mechanism of this loss is still a subject of heated debate; some have argued that the store is subject to a rapid time-based decay unless it is rehearsed (Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007; Portrat, Barrouillet, & Camos, 2008), whereas others have argued that the loss of information is due to the interference during the delay (Lewandowsky & Oberauer, 2009).

Evidence for the subvocal rehearsal process is provided by the word length effect, that is, the memory span declines with lists composed of lengthier and multisyllabic words (Baddeley, Thomson, & Buchanan, 1975). This phenomenon appears to arise from the greater decay of phonological representations in the store caused by the longer time taken to subvocally rehearse lengthier items. Rehearsal is prevented when participants engage in an activity known as articulatory suppression, in which irrelevant information is continuously repeated, thereby eliminating the word length effect (Baddeley, et al., 1975; Murray, 1968). When the rehearsal content exceeds the capacity of the phonological loop, the rehearsal strategy can be strategically abandoned by participants (Salamé & Baddeley, 1986).

The capacity of the phonological loop is typically measured by immediate serial recall, such as recalling an unfamiliar sequence of digits or repeating a sentence. Several computational models have been constructed to account for the mechanism of the serial rehearsal. The primacy model assumes a primacy gradient of activation of successive items, such that items earlier in the list are more active than later ones. This is followed by a repeated cycle of a noisy item choice and later a suppression of the chosen item (Page & Norris, 1998). The start-end model, however, suggests that the coding of the item position is based on the start as well as the end of the corresponding sequence, and the relative position is used as the cue for recall (Henson, 1998). In contrast to emphasizing the importance of ordinal cues in processing serial verbal materials, other models consider verbal representations to be multidimensional. The feature model postulates that items are represented as vectors of features, which can be selectively overwritten by subsequent external events and also by the ongoing stream of internal activity (Nairne, 1990; Neath & Nairne, 1995). In a more complex contextual-based model, each item is represented as multiple layers, including lexical, timing, input and output phonemic information, with a context vector representing the serial position. Recall is realized by rerunning the time signal and reactivating each positional context vector in order, resulting in a sequence of most activated item as time evolves (Burgess & Hitch, 1992, 1999). However, none of the aforementioned models can account for all the characteristics of rehearsal process, implying the complex nature of the processing of serially-presented verbal materials.

The visuospatial sketchpad

The visuospatial sketchpad specializes in the maintenance of visual and spatial information (Smyth, Pearson, & Pendleton, 1988; Smyth & Pendleton, 1990), and has a capacity of three or four objects in adults (Baddeley, 2003b; Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). According to Logie (1995; 2003), the visuospatial sketchpad includes a passive storage system 'visual cache', similar to the phonological store's retaining of the visual properties of objects or scenes, and an active spatially-based rehearsal mechanism acting as an 'inner scribe' to support the planning and cognitive control of movements.

The visual store is able to hold up to four objects containing multiple features (Luck & Vogel, 1997), and attending to one feature of the object can automatically activate the other features (Egly, Driver, & Rafal, 1994). Moreover, this store is resistant to interference decay and appears to store serial order (Baddeley, 2007; Logie, Sala, Wynn, & Baddeley, 2000).

The underlying spatial rehearsal mechanism is less clear. According to Logie (1995), active rehearsal occurs mainly in the course of encoding spatial information, which involves the shift of spatial attention (Awh, Jonides, & Renter-Lorenz, 1998; Smyth & Scholey, 1994). In addition to attentional shift, eye movement was found to be involved in remembering spatial locations, in contrast to the lack of the role played by this in the storage of visual information. Moreover, it was the cognitive control of eye movement rather than the movement itself which was discovered to be underlying the encoding, maintenance and retrieval of spatial information (Postle, Idzikowski, Della Sala, Logie, & Baddeley, 2006). Nevertheless, interference from eye movement was found to be larger than from attention shift to a spatial working memory task, implying that a mechanism other than shifting attention contributes to the spatial rehearsal. It has been speculated that this additional disruption was caused by the change of coordinates of to-be-remembered locations during eye movement and the cognitive suppression of spatial processing during the execution of eye movement (Lawrence, Myerson, & Abrams, 2004). The disruptive effect of movement on spatial working memory performance is not restricted to eye movement, but also shows in other types of movement, such as pointing and arm movements, indicating that there is a connection between movement and spatial working memory (Hale, Myerson, Rhee, Weiss, & Abrams, 1996; Lawrence, Myerson, Oonk, & Abrams, 2001; Quinn & Ralston, 1986). Nevertheless, there is also evidence suggesting an independent motor component (Smyth, et al., 1988; Smyth & Pendleton, 1989).

This separation of visual and spatial subsystems in the visuospatial sketchpad mirrors the 'what and where' organization in the visual system (Mishkin, Ungerleider, & Macko, 1983), and is supported by several behavioural experiments (Della Sala, Gray, Baddeley, Allemano, & Wilson, 1999; Klauer & Zhao, 2004; Tresch, Sinnamon, & Seamon, 1993) as well as by the double dissociation found in brain-damaged patients (Della Sala, et al., 1999; Farah, Hammond, Levine, & Calvanio, 1988; Luzzatti, Vecchi, Agazzi, & Vergani, 1998) and the neuroimaging evidence (Baker, Frith, Frackowiak, & Dolan, 1996; Hautzel, et al., 2002; McCarthy, 1996; Smith, et al., 1995). However, a recent review in brain research showed that, although there was a clear dorsal-ventral distinction in maintenance of objects and locations, processing spatial and visual information was found to employ similar parts of the prefrontal cortex (Wager & Smith, 2003a).

Given the above literature, the visual, spatial and movement subcomponents in the visuospatial sketchpad can be said to be both separable and interactive; because of this, pure interference tasks that selectively disrupt these subcomponents have proved challenging to develop. One relatively pure measure of visual short-term memory is the visual pattern task, which requires the immediate reproduction of a partially filled-in matrix (Della Sala, et al., 1999). Interference tasks used in studies of visual short-term memory include watching irrelevant pictures (Logie, 1986) or abstract patterns (Quinn & McConnell, 2006). Memorizing sequential spatial locations is often measured by the Corsi-block task, that require memorizing the order of a set of blocks being tapped (Berch, Krikorian, & Huha, 1998; Corsi, 1972). The tapping task is a commonly-used interference task to disrupt the spatial component in the visuospatial sketchpad. It usually involves movement to external spatial targets, such as repetitive tapping according to a predetermined sequence by moving hands to different locations in space (Farmer, Berman, & Fletcher, 1986; Smyth, et al., 1988). A motor interference task can either involve body-related movement, like repeating a sequence of body movements, such as touching the top of the head, then the shoulders (Smyth, et al., 1988), or involve objects, for example, squeezing and releasing a soft tube held in the hand when bending the arm towards the body (Rumiati & Tessari, 2002).

The central executive

The central executive is the attentional system that regulates the two storage systems (phonological loop and visuospatial sketchpad), and also retrieves information from the long-term memory into the working memory (Baddeley, 2007). It is said to have a range of executive functions, including switching task and strategy, updating, inhibiting, as well as focusing, dividing and switching attention (Baddeley, 1996, 2007). These functions are closely related to executive functions, an umbrella term for a wide range of cognitive processes and behavioural competencies like planning, sequencing, and monitoring actions (Chan, Shum, Toulopoulou, & Chen, 2008).Both central executive and executive functions are complex concepts and the main debate hinges on its unity or diversity. Findings by Miyake and his colleagues (2000) indicated that three executive functions (shifting, updating and inhibition) are moderately correlated with one another but also separable, supporting both Baddeley's attempt to fractionate the central executive (Baddeley, 1996) and the idea of a unitary factor earlier (Stuss, Shallice, Alexander, & Picton, 1995b). According to Engle and his colleagues, the commonality between executive functions and central executive is reflected in the attention control via the active maintenance of goals and the inhibition of irrelevant information (Engle, Kane, & Tuholski, 1999; Engle, Tuholski, Laughlin, & Conway, 1999). The difference between the subcomponents of the central executive is reflected in the neuroimaging evidence that different brain areas are activated for different functions. Specifically, mental operations, switching, and inhibition activate the inferior frontal cortex, whereas continuous updating, sequential organizing, and prioritizing information activate the superior frontal cortex (Wager & Smith, 2003a).

There are several tasks that selectively interrupt the central executive. One is the random generation task, which requires producing a random sequence of letters or pressing an array of ten keys at random at a varied production rate; the faster the production rate, the greater the demands on the executive (Baddeley, 1966). This task requires the inhibition of natural retrieval strategies and a search for alternative ways to generate letters, thus involves subcomponents of central executive such as updating and inhibition (Miyake, et al., 2000). Another task, called the backward counting task, requires continuous deduction of one or two from a three-digit number (Postma & De Haan, 1996). The load of central executive is manipulated and reflected as the size of the subtrahend. The backward counting task requires retrieving the subtraction rules from the long-term memory and applying them to perform the arithmetic operation, therefore drawing upon the central executive. Moreover, people need to hold the intermediate products which tend to use the phonological loop (Seitz & Schumann-Hengsteler, 2002). Both the random generation task and the backward counting task impose on one of the storage components, either the phonological loop or the visuospatial sketchpad. A task that draws on the central executive without additional storage demand is the random interval repetition tapping task. In this task, participants react to sounds that occur at random time points as quickly as possible; this requires constant attention and monitoring, and therefore loads the central executive resources (Vandierendonck, De Vooght, & van der Goten, 1998).

The episodic buffer

A fourth component, the episodic buffer, was added more recently (Baddeley, 2000). This is a multi-modal temporary storage system, capable of binding information from the other components of working memory, as well as from the long-term memory and various perceptual channels, into a single coherent episode. Its capacity is assumed to be limited by the number of chunks or episodes (Baddeley, Allen, & Hitch, 2011).

Three lines of research have investigated the question of binding, including the binding of visual features (Allen, Baddeley, & Hitch, 2006; Brown & Brockmole, 2010; Ueno, Mate, Allen, Hitch, & Baddeley, 2011), the binding of words (Baddeley, Hitch, & Allen, 2009; Jefferies, Ralph, & Baddeley, 2004) and cross-modal binding (Allen, Hitch, & Baddeley, 2009).

In the visual domains, the process of binding is still controversial. There is evidence supporting the automatic binding of features (Luck & Vogel, 1997; Vogel, et al., 2001) in contrast to the attention demanding view held by Wheeler and Treisman (2002). Baddeley et al. (2011) reconciled the two views by suggesting the automatic binding in the initial stage, and an attentional control to prevent disruption from competing stimuli in order to maintain the visual features. The argument for automatic binding is supported by the findings that recognition of combined features and single features were disrupted by a concurrent task tapping on the central executive to a similar degree, suggesting that binding features requires little attentional control and occurs automatically (Allen, et al., 2006). Moreover, the automaticity extends to both temporally and spatially separate features (Karlsen, Allen, Baddeley, & Hitch, 2010). However, one study found that the central executive was significantly involved in binding colours and shapes (Brown & Brockmole, 2010).

Although encoding bound features appears to be relatively automatic, their maintenance seems not to be. In an experiment comparing serial and simultaneous presentation of bound and single features, bound features suffered more than single features in serial presentation, suggesting that the maintenance of bound features is relatively fragile. Moreover, this difference in the recognition accuracy between bound and individual features was much larger in the early serial positions than in the later ones, suggesting that the maintenance was not only fragile but probably susceptible to interference (Allen, et al., 2006, Experiment 5). The fragility of holding bound features in memory was explored in a series of experiments using the suffix paradigm, in which a to-be-ignored suffix was added to the end of to-be-remembered features (Ueno, Allen, Baddeley, Hitch, & Saito, 2011). They found that when the features of the suffix overlapped with the feature pool of the to-be-remembered features, it created interference and impaired recognition of the bound features more than of the single features, suggesting that bound features are more fragile and susceptible to interference. This fragility was not caused by the increased attentional demands of filtering out similar suffixes, because the two-feature-overlap suffix and the one-feature-overlap suffix had similarly disruptive effects on the memory of bound features (Ueno, Mate, et al., 2011).

Another line of research has explored binding within sentences, that is, whether the central executive contributes to the benefit of chunking in the sentences. In one study (Jefferies, et al., 2004), a concurrent attentional demanding task (the choice reaction time task) was found to disrupt the recall of auditory unrelated sentences not that of a story or unrelated word lists nor unrelated word lists. The authors thus inferred that attention is involved in forming links between unrelated propositions, whereas syntactic and semantic factors operate relatively automatically. A later study, however, found a similar level of involvement of the central executive in constrained sentences, word lists and open sentences, suggesting that the central executive plays no special role in chunking (Baddeley, et al., 2009). In the former study, central executive was found essential in binding unrelated short sentences, whereas in the latter study, binding constituents within a constrained sentence was found relatively automatic. It seems, therefore, that binding is likely to occur at a higher level of organization rather than at a lower level. This also corresponds to the proposal of Baddeley et al. (2009) that the episodic buffer is a relatively passive episodic storage and it is the operation outside the buffer requires executive processing.

Besides binding within a domain, the episodic buffer is also assumed to combine information from various modalities. However, a recent study showed that cross-modal combination of visual and auditory information did not demand more of central executive resources than unified colourful shapes, suggesting the automaticity of the cross-modal binding in forming a visual image (Allen, et al., 2009). Nevertheless, the role played by the central executive in binding of cross-modal information to form a concept or sound image remains unknown. On the basis of the aforementioned findings, the episodic buffer has been revised to be a purely passive 'screen' that is fed by information from subsystems of the working memory (phonological loop and visuospatial sketchpad) and also from the long-term memory, forming a multidimensional episode that is available to the conscious awareness (Baddeley, Allen, & Hitch, 2010; Baddeley, et al., 2011).

Separating working memory components

One important methodology for teasing apart working memory components is the dual task methodology (Baddeley, 1986), employing the logic that tasks using the same cognitive components will compete for resource hence simultaneous processing of two tasks will lead to a decrement of performance; in contrast, tasks using different components will not. Another way of investigating these components is through latent variable analysis exploring the relationships between them (Alloway, Gathercole, & Pickering, 2006; Kane, et al., 2004; Park, et al., 2002).

The distinction between storage and manipulation of information is reflected in the tasks being used. Short-term memory tasks require only that information is held for a short period of time, whereas working memory tasks require both storage and manipulation of information. Research also supports the notion that domain-specific storage and rehearsal relate more strongly to domain-specific complex cognitive activities, whereas working memory is a stronger predictor of general fluid intelligence (Kane, et al., 2004). This working memory structure is relatively stable throughout the human life span, from young children to old adults (Alloway, et al., 2006; Park, et al., 2002).

It is worth noting that there is evidence suggesting a less separable relationship between the visuospatial sketchpad and the central executive. Studies using dual task methodology showed that memory for visual patterns was disrupted by an auditory mental arithmetic task requiring the central executive (Phillips & Christie, 1977); visualizing of spatial routes in the Brooks task experienced significant interference from a concurrent executive demanding task (the random generation task) (Salway & Logie, 1995). A study using structural equation modelling also found that storage-oriented visuospatial short-term memory tasks tended to involve aspects of central executive functioning (Miyake, Friedman, Rettinger, Shah, & Hegarty, 2001). One speculation as to the reason for the involvement of this appears related to sequencing. Several studies have shown that the central executive tends to be used when the visuospatial tasks require the retention of sequential information (Fisk & Sharp, 2003; Klauer & Stegmaier, 1997). This led Jones et al. to emphasize the difference between spatial tasks requiring preservation of order information and those requiring only maintenance of the pattern information (Jones, Farrand, Stuart, & Morris, 1995b). This viewpoint has gained support from a recent study in which the performance of the visuospatial tasks that involved sequential processing was more impaired by an executive demanding task (random digit generation) than were visuospatial tasks that involved simultaneous processing (Rudkin, Pearson, & Logie, 2007). The authors explain the increasing demand of the central executive as being due to the increased strategic control needed to actively construct mental path configurations during sequential encoding. However, others have suggested that it is due to the shift and selection of attention underlying the maintenance of sequentially presented spatial items (Awh, et al., 1998; Smyth & Scholey, 1994).

The separation of the two storage systems, the phonological loop and the visuospatial sketchpad is demonstrated by the classic Brooks task (Brooks, 1967), in which participants recall either a verbal description of a designated sequence requiring putting digits in a 4×4 empty matrix (spatial task), or a series of similar but nonsensical sentences (verbal task). The former was found to be selectively disrupted by a secondary visuospatial task but less so by a verbal task; and the reverse was observed in the verbal version of the Brooks task. This suggests that spatial imagery depends on the visuospatial sketchpad, whereas remembering verbal sentences requires the phonological loop (Salway & Logie, 1995). Other evidence supporting the separation includes double dissociation in neuropsychological patients (Della Sala & Logie, 2002; Vallar & Papagno, 2002), as well as imaging research in which a verbal working memory task has been found to use the left hemisphere (supplementary and premotor areas) whereas the visuospatial working memory task has been shown to involve mainly the right hemisphere (premotor and dorsal lateral prefrontal lobe) (Smith & Jonides, 1999; Smith, Jonides, & Koeppe, 1996; Wager & Smith, 2003a).

Summary

Three working memory models have been evaluated in order to choose an appropriate one that serves the purpose of current research, to explore the functioning of working memory in following instructions. Atkinson and Shiffrin's modal model (Atkinson & Shiffrin, 1968) does not make a distinction between maintenance and manipulation; for this reason it is inadequate in help representing a complex activity like following an instruction, in which recoding the instruction requires both maintenance and manipulation of information.

Cowan's model (1999) focuses on the interaction between attention and memory activation. Similarly to the multicomponent working memory model, it admits a distinction between an active processing component and a passive storage component. However, the model holds a unitary view in the sense that, notwithstanding the existing of different types of domain-specific representations, the same rule of processing (memory activation) is followed. Importantly, working memory is considered to be an active process that summons all possible mechanisms and resources in order to complete a particular task. Nonetheless, it does not specify the way in which a complex task is accomplished with each of these mechanisms; rather, it emphasizes an overall capacity limit on cognitive performance. Therefore, in order for the functional organization of cognition in a complex task to be understood, models which emphasize type rather than the amount of cognitive processing seem to be more suitable for this purpose (Logie, 2011).

Therefore, the multi-component working memory model (Baddeley, 2000) seems to be the one that suits this purpose best. It is theoretically valuable in providing explanations for a variety of complex cognitive activities, especially for learning new materials. Moreover, given the limitations of working memory capacity, people are required to manoeuvre their working memory resources in order to form the most useful representation for a cognitive demanding task. Thus, the multicomponent working memory model also allows an investigation of people's strategic reliance on specific working memory components in the process of following instructions.

Another benefit of this model is that it introduces the concept of the episodic buffer, which is in charge of combining information from distinct processing resources into a coherent episodic chunk for future use. This function is likely to be highly useful in encoding instructions, because most commands are made up of elements from various sources that needed to be encoded into a coherent action episode to achieve the action goal. Finally, the multicomponent model provides a well-developed technique, namely, the dual task methodology. This method is useful for separating contributions from different working memory components, and was thus suitable for the purpose of current research.

Notwithstanding the many merits of the multicomponent working memory model, it places less emphasis on the exact role of central executive and the role of long-term memory. Therefore, the alternative view that emphasizes the attentional control of memory activation should also be considered. Together, these models may provide a comprehensive view on the cognitive process involved in a complex task like following instructions.

The interface between perception, memory, action and language

Following instructions is a complex task that involves multiple cognitive functions, such as perception, language, memory, and action. For example, remembering and understanding oral commands involves speech perception, while decoding written instructions demands reading skills, and both may need rehearsal for temporary storage of the information. Demonstrations of actions often contain information from multiple media, and the formation of a coherent action episode is likely to draw upon the episodic buffer. More importantly, these cognitive functions interact with each other, that is, they work together to complete the task. As soon as instructions begin to be encoded, a dynamic relationship develops between language, perception and memory. For instance, action commands often involve operations upon objects scattered in space, and people tend to track these to-be-enacted objects as the names of these are mentioned, which requires intimate cooperation between language perception, comprehension, memory and action planning. This section reviews literature on the interactions between perception, language, memory and action (Henderson & Ferreira, 2004). I will first introduce the ventral and dorsal systems in perception and action, and then discuss the conjoined work of the relevant cognitive functions in perceiving commands and planning actions, as well as retrieving the plans of actions during recall.

Ventral and dorsal system in perception and action

Based on findings in neuropsychological and imaging research, Milner and Goodale argued that the dissociable process in perception and action was due to their different purposes (2006b). The perception-aimed process involves the ventral stream in the human brain, and is aimed at identifying, classifying and attaching meanings to objects for later responses. Therefore, the enduring properties, such as the texture gradients, colours and shadings, are all important in perception. In contrast, action-directed processes involve the visuomotor system, and use the dorsal pathway in the brain. They handle the moment-to-moment practical problems, like operations upon objects; locations and motions are thus crucial characteristics in this process. It seems, then, that the dorsal stream deals with viewer-centred coding in order to control object-directed action while the ventral stream forms more generalized representations.

Although different, under the direction of selective attention, the two systems can work together to achieve the goal of an action (Milner & Goodale, 2006a). The facilitating role of visual attention in visually-guided actions is reflected in the shared frontal-parietal circuits in the brain (Rizzolatti, Riggio, & Sheliga, 1994). According to Milner and Goodale (2006a), in both streams visual attention is devoted to intended objects, which are 'flagged' to achieve certain ends, conscious perception in ventral stream and action for the dorsal stream. Visual attention is also driven by important visual information that bears great evolutionary significance, such as motion, which is processed by both streams.

Objects often act as goals to facilitate actions (Gattis, et al., 2002; Vogt, 2002). Neurophysiological evidence indicates a polysensory brain area that receives inputs of object recognition in the ventral stream, and spatial and action information in the dorsal stream. For example, when an action-related goal is created, such as generating associated action words, it does not matter whether participants see a picture or the name of an object; both of these activate complex neural networks relating to the semantic aspects associated with knowledge of motion as well as the grammatical aspects, i.e. the functions of verbs (Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995). Thus, it appears that the goal of an action dictates the way in which the two systems collaborate.

Cooperative work of cognitive functions in following instructions

The first step in following instructions is to understand the commands given, and language comprehension is supported both by working memory and semantic long-term memory (Baddeley, 2003a; Daneman & Merikle, 1996; Jefferies, et al., 2004). After comprehending the instructions, next step is to translate abstract linguistic codes into actionable commands.

According to Paivio and Koriat (Koriat, et al., 1990; Paivio, 1986a), action-oriented contents are likely to be stored in an imagery-motor-based representation that is closely linked to the external world. The process of mapping linguistic input onto an action-based representation starts from the earliest moment of processing, reflected as dynamically updated referential domains (Tanenhaus, Chambers, & Hanna, 2004). One way of mapping this is reflected in the time-locked relation between eye movement and speech. For example, while listening to instructions containing future actions upon objects, people tend to shift their gazes to the objects once they recognize the spoken words referring to them (Griffin, 2004). People can identify objects in less than 170ms (Potter, 1975) and tend to gaze at the objects until they have retrieved the phonological form of their names (Meyer, Sleiderink, & Levelt, 1998).

According to Spivey, looking at objects and remembering their locations is an efficient way of using the external environment to encode overwhelming visual details in a three-dimensional world (Spivey, Richardson, & Fitneva, 2004). The process involves creating deictic pointers, addresses of object locations in the environment, along with labelling information about when and why to use these pointers. It has been found that as soon as encoding begins, eye movements mirror the spatial information in a spatiotemporally dynamic scene (Spivey & Geng, 2001).

Besides being responsive, eye movements can also be predictive. Mediated by language, these anticipatory eye movements can direct towards an object to be mentioned, which can be inferred based on the linguistic input status quo (Altmann & Kamide, 2004).

Sometimes, actions have to be postponed until the proper time point. Actions based on memory are prone to errors; for example, memory-driven saccades towards objects, and the process of grasping these objects, are generally inaccurate and slower. For example, with the imposition of a two-second delay, the pre-shaping of the hand becomes less accurate, and the path followed by the hand is more curvilinear than hand movements in situations with no delays (Goodale, Jakobson, & Keillor, 1994; Milner & Goodale, 2006b). To ensure the accuracy of actions, a relatively late action planning seems better than an early one when the action has to be delayed; that is, a detailed plan for action may not be programmed until the moment it is enacted in order to avoid perturbations during the delay (Westwood & Goodale, 2003). It is possible that, in the situations when delay is long, it is the intention of actions that is maintained rather than the layout of the action plan. However, the hypothesis of late action planning is still the subject of much debate.

The cognitive process of executing actions has been reviewed in the previous section with an emphasis on the monitoring process. I will now focus on situations in which intended actions are required to be recalled verbally. The mapping of an unordered multidimensional conceptual content onto a grammatically ordered sequence of words is considered difficult (Wundt, cited in Griffin, 2004). The eye movements to referents are thus helpful in the way that they provide converging kinaesthetic and spatial order cues for message elements, which can help decrease sequencing errors. In a situation where speakers retrieve a message linguistically, the message element also activates its associative spatial index, triggering eye movements to that region of space (Griffin, 2004). It could be argued that gazing is helpful because the visual forms can be used as external semantic representations to guide the word production. However, when an object is no longer in the place where it was originally being stored, people still tend to look at that place. Therefore, it seems gaze is driven to locations rather than objects per se, reflecting an automatic attempt by an embodied working memory system to access the contents of a spatial pointer's address in an external environment (Spivey & Geng, 2001).

Summary

As can be seen, translating instructions into actions is complicated, involving multiple cognitive functions and also relying on the intimate cooperation of these. It requires integrating the commands of others into one's own mental representations, and then mapping them back onto an external world. Three points are worth emphasizing here. First, among the many cognitive functions that have contributed to this process, it is important to highlight the crucial role of visual attention. From the very start of encoding, voluntary eye movement begins to build the links between commanding codes and to-be-enacted objects. The deployment of a spatial representation of actions eases the process of encoding as well as the process of retrieval, during which the eye movement again has a role.

Second, although actions are often the ultimate goals and ends of an instruction, the role of language in supporting the construction of the representations of actions should not be overlooked. Moreover, during execution, control of action may be assisted by the subvocalization in order for the actor to maintain strategic control of his or her performance (Baddeley, 2003b; Baddeley, et al., 2001). Sometimes, repeating or rephrasing the instructions can be the aim and endpoint, such as the circumstance of a teacher giving oral orders in a classroom. Therefore, language serves as a carrier as well as a mediator for giving orders and guiding actions.

Finally, following instructions requires encoding, maintaining information and monitoring execution of actions, which is especially pertinent to working memory, an ability to maintain and manipulate information in a short period of time. Moreover, working memory is likely to be critical for interweaving various relevant cognitive functions and processes into a coherent representation, maintaining it, and monitoring it until the goal is achieved. Therefore, the next section will take a closer look at studies that exploring the involvement of working memory in following instructions.

Working memory and following instructions

Correlational studies

Since the work of Binet and Thorndike, the ability to follow instructions is considered to be a measure of intelligence, and this ability to execute a series of actions increases with age (Binet & Simon, 1912; Thorndike, 1927, cited in Kaplan & White, 1980). Early investigations like that of Brener (1940) gave participants simple commands like 'put a comma below', presented through an exposure apparatus at the rate of two seconds per command. The length of commands varied from one to five, and participants performed the task in sequence on a card using pencil. The mean span of the university students was 2.42. Moreover, Brener noticed that the ability to follow instruction was significantly correlated with the digit span, a measure of short-term memory (Brener, 1940).

Later studies focused on grammatical complexity. For example, in the Token test, participants were required to carry out instructions, such as 'after picking up the green rectangle, touch the white circle', while the length and grammatical complexity of the instructions was systematically varied (De Renzi & Vignolo, 1962). In this task, years of schooling, but not age, significantly affect the performance (De Renzi & Faglioni, 1978). Importantly, the Token test was found to be significantly correlated with the verbal, visual and motor aspects of short-term memory (Lesser, 1976). The Token test was mostly used to discriminate subtle oral comprehension difficulties in adult aphasic people (De Renzi & Vignolo, 1962) and was later adapted for both normal children and children with language delay (Cole & Fewell, 1983), as well as being used in a clinical paediatric population (Paquier, et al., 2009).

The interest of classroom instructions began with Kaplan and White (1980). They analyzed teachers' classroom instructions in elementary school and located two sources of complexity of instruction; one is grammatical, specifically relating to qualifiers (e.g. who, where, when and how), and the other is the number of behaviours. Thus, a direction such as ‘open your books to page three and do the first three problems’ contains two behaviours and two qualifiers. The instructions were administrated to 215 children from grades K-5. Children were required to execute these instructions immediately after hearing they hear them read aloud. It was found that the ability to follow instructions steadily increased over grades K-3 and levelled off in grades 4-5, which may be due to a ceiling effect. Moreover, increasing the sentence complexity (by adding qualifiers) impaired the performance, especially for children in grades K-2. Although Kaplan and White did not give a memory test to children, it would seem likely that the increasing sentence complexity reflects an increasing demand on working memory.

More direct proof of the relationship between working memory and following instruction came from the study of Engle and his colleagues (Engle, et al., 1991). They adapted Kaplan and White’s instruction task (1980) and included both a pencil-and-paper task (e.g., ‘point to the picture at the top of page three and copy it twice’) and action-oriented task (e.g., ‘sit on the floor Indian-style’). Consistent with the findings in the former study (Kaplan & White, 1980), there was significant performance improvement in children aged 7, 9 and 12. Memory storage (measured by word span) and working memory (measured by sentence span) was found to have a close and also increasing relationship with instruction performance as age increases. Moreover, compared to children with a high working memory span, children with a low working memory span had more difficulty in following more complex instructions than simpler instructions.

Following Engle et al.’s study (1991) and based upon observation in the classroom and a pilot study, Gathercole and Alloway (2008) noted that children who score poorly on central executive measures have marked difficulties in carrying out instructions in the course of their everyday classroom lives. To examine this more specifically, a classroom instruction task was designed (Gathercole, et al., 2008). The instructions varied only in length of steps (behaviours) and were matched in both grammatical complexity and number of behaviours to exclude the language development confounding. Five-year-old children listened to instructions like ‘Touch the red pencil, then pick up the blue ruler and put it in the black box, then pick up the white eraser’, and were required to recall this either by repeating the instruction sentence or carrying out the actions upon the colourful stationery. Children were found to be better at performing instructions than repeating them, and the accuracy of performing but not repeating instructions was strongly associated with working memory, including both storage (measured by the forward digit recall task) and the processing ability (measured by the backward digit recall task). Moreover, the association between following instruction performance and manipulation was found to be stronger than simple storage. The superiority of action recall was explained as being due to the benefit of the motoric or spatial integration in the encoding stage in contrast to the verbal representation assumed to be sufficient for oral repetition.

The close relationship between working memory and the ability to follow instructions was found to exist not only in children, but was also observed in young and older adults (Kim, Bayles, & Beeson, 2008). In their study, instructions varied in both length of actions and complexity of qualifiers, and were adapted to familiar daily situations experienced by older adults (e.g. ‘Take two red pills on Saturday morning’). Participants responded by putting pills into a compartment representing the date. Both short-term memory (measured by digit span) and age were significant predictors of instruction performance. Moreover, participants performed more accurately when the instruction contained fewer actions, even if the action contained more qualifiers, suggesting that it was the length of the actions that mattered. This is different from the results in children, in which the qualifiers were an important indicator of performance in following instructions (Kaplan & White, 1980).

Not only in children, the close relationship between working memory and the ability to follow instructions was also found in young and older adults (Kim, et al., 2008). In their study, instructions varied in both length of actions and complexity of qualifiers, and adapted to older adults' daily situation (e.g., Take 2 red pills on Saturday morning). Participants responded by putting pills into compartment representing the date. Both short-term memory (measure by digit span) and age were significant predictors of instruction performance. Moreover, participants performed more accurately when the instruction contained fewer actions, even the action contained more qualifiers, suggesting it was the length of actions that matters. This is different from the results in children, in which the qualifiers are important indicator of performance of following instructions (Kaplan & White, 1980).

Experimental studies

Although there were some discrepancies between these limited studies on following instructions, they all implied that there is a close relationship between working memory and following instructions. More direct evidence came from a recent exploration by Allen (2009) using laminated cards of geometric shapes (e.g. star, cross). Instructions were read out by the experimenter, a typical one being 'Push the cross, spin the star, drag the arch, and touch the square'. Adults either repeated back the instruction sentence or performed on the laminated cards by hand. As in the experiment with children (Gathercole, et al., 2008), performing the actions was found to be superior to the spoken repetition.

Moreover, the involvement of working memory components in following spoken commands was investigated using the dual-task methodology under the theoretic framework of multicomponent working memory model (Baddeley & Hitch, 1974). Allen (2009) found that performance of recall was impaired both by articulatory suppression and backward counting, suggesting significant contribution of both phonological loop and central executive. To explore the visual-spatial sketchpad, participants were instructed to look away from the visual display of the laminated cards during instructions, which blocked their access to the visual and spatial information. The disruptive effect was evident only in performance of recall by actions but not in performance of oral repetition, but this result was not replicated in a later experiment, thus leaving the contribution of the visuospatial sketchpad unclear.

Sequential representations

Correctly recalling or performing instructions requires not only remembering the correct actions, but also recalling them in the correct sequence. This is especially true for a relatively new sequence of actions in which cause-effect links between steps are obscure, unlike a familiar sequence of actions that may be supported by motor schemas from the long-term memory. Indeed, sequential representations instantiate in multiple facets in a task such as following instructions. For example, spoken commands are comprised of a series of words containing a flow of phonemes. Remembering the action sequence requires representing them in a chain of actions for later execution. An action such as grasping also contains a series of well-learnt movements, while the oral repetition of instructions requires organizing words in a sequential manner.

A question following on from this is how sequential information is processed and represented. Given the omnipresence of sequences in perception, speech, motor control, one might expect there to be a specialized system for processing serial order information. Research has shown that there are indeed many similarities between sequences in different domains, such as the many similarities found in the verbal and visuospatial sequences (Hurlstone, 2010). Interference studies also indicate that when the secondary task includes an order component or a changing state, the memory of the primary serial memory task is impaired, irrespective of the modalities (Depoorter & Vandierendonck, 2009; Jones, et al., 1995b). In contrast, the multicomponent working memory holds the assumption that order information is maintained separately within domains (Baddeley, 2007). For example, a serial verbal task does not interfere with spatial recall, and serial spatial tapping does not interfere with recalling sequences of body movement (Smyth, et al., 1988).

This puzzle can perhaps be explained by the existence of two sequence systems which have been suggested by Keele and his colleagues: a unidimensional system that specializes in processing information in a single dimension, and a multidimensional system that builds associations between events from different dimensions or modality domains (Keele, Ivry, Mayr, & Hazeltine, 2003). In the unidimensional system learning is implicit and occurs automatically; moreover, it is not susceptible to potentially disruptive information from other dimensions, which explains the lack of interference from the sequence of a different dimension.

The merging of sequential information from various dimensions is actualized by the selective attention directed by the goal of learning. The predictability among events from various channels promotes interdimensional learning, whereas randomness discourages learning. The multiple representations of information may sometimes be considered redundant; nevertheless, this redundancy provides additional contextual information that helps decrease the ambiguity arising from difficult sequences. The co-existence of the two systems increases the flexibility of processing ordinal information, as well as highlighting the benefit of cross-modal representations when learning in a complex environment.

Although the nature of the sequence representation is still contentious, several robust effects in a typical serial recall have helped illuminate the cognitive process of encoding and retrieving the sequential information. A typical serial recall curve is a bow-shaped curve with a high percentage of recall from the beginning and end of a sequence, known respectively as the primacy and recency effects. The primacy effect occurs because the items in the beginning of a sequence are more distinctive, as there are far fewer items being processed at that time (Glenberg & Swanson, 1986). Since people tend to rehearse a sequence from the beginning, the initial items also benefit from the repetitive rehearsal that facilitates transference to the long-term memory (Burgess & Hitch, 1992). The greater distinctiveness and more rehearsal time devoted to these initial items mean that, during the encoding of a list or sequence, less and less attention is devoted to additional list time, known as the primacy gradient (Page & Norris, 1998).In the recall stage, the items retrieved later tend to receive output interference from earlier items, resulting in a primacy gradient of items (Cowan, Saults, Elliott, & Moreno, 2002; Oberauer, 2003). Sometimes, forgetting the first item may lead to the loss of the entire action sequence, such as playing a piece of music, when it has to start from a particular section. Similarly, in situations of following instructions, remembering the first step may also be crucial in anchoring the starting position that helps initiate the following steps.

The recency effect, on the other hand, is a result of the lingering presence of the last items in the working memory when recall is required. Therefore, when the time to recall is delayed, or the content of recall is too long, the recency effect is likely to be reduced. The recency effect can also be explained as retroactive interference during encoding; that is, the earlier items suffer more interference from the later ones in the list (Nairne, 1988). Alternatively, it could be due to response suppression, which is when an item is removed from memory once it is recalled to avoid its preservation. Consequently, the later items being recalled have less competition from the earlier items, and therefore the probability of recalling the final items correctly is increased, thus generating a recency effect (Farrell & Lewandowsky, 2002). In addition, the last item usually marks the end of a task hence probably worth being remembered. In addition, the last item usually marks the end of a task, and hence is probably worth being remembered. Finally, it is worth mentioning that the extent and magnitude of the recency effect vary with modalities; it is usually larger in aurally-presented sequences than in those which are visually presented. This modality effect could be either due to the benefit of echoic persistence of sound after its physical stimulation is ceased (Watkins & Watkins, 1980), or to the superior temporal representation of information presented aurally rather than visually (Glenberg & Swanson, 1986).

Errors are common in serial memory, including both item and order errors. Common item errors include repetitions (where an item is recalled more times than it was actually presented), omissions (where an item is not recalled) and intrusions (where an item is recalled that was not presented). Order errors include anticipation errors (when an item is recalled ahead of its position), postponement errors (where an item is recalled at a later point in the sequence than its correct position), and exchange errors (where two items swap positions) (Hurlstone, 2010). Interestingly, people tend to commit more order errors than item errors; in fact, order errors account for around 80% of total errors (Aaronson, 1968). This again implies the difficulty of retaining sequential information.

Overview of the thesis

Findings and research questions

Following instructions is a common activity that supports learning in everyday life. From this review of a wide number of studies relating to following instructions, it is clear that the underlying cognitive process is very complex. Importantly, studies have indicated the importance of working memory in following instructions (Allen, 2009; Engle, et al., 1991; Gathercole, et al., 2008; Kim, et al., 2008). As yet, little is known about the specific roles played by subcomponents of working memory. Understanding the differential roles of these subcomponents may help clarify the cognitive process of following instructions. Therefore, the first aim of this thesis is to investigate the contribution of working memory components under the multi-component working memory model proposed by Baddeley and Hitch (1974).

The original version of the working memory model was used to guide this research (Baddeley & Hitch, 1974). This contains three components: two storage systems, the phonological loop and the visuospatial sketchpad, and a supervisory system, the central executive. Both storage systems are expected to be involved in the successful following of commands. The phonological loop serves to rehearse the content of instructions and put them in its store. The visuospatial sketchpad functions to search relevant objects and store the information of movement. The central executive is thought of as exerting conscious control during the planning and execution of actions.

Another noteworthy phenomenon is the superiority of action response compared to verbal response, which is reflected in greater accuracy and fewer errors (Koriat, et al., 1990), as well as in error corrections (Prinz, 2002). This benefit of recalling by actions than repetitions is also present in studies of following instructions, in which recalling instructions by execution leads to superior performance of recall compared to simply repeating the instructions verbally (Allen, 2009; Gathercole, et al., 2008; Koriat, et al., 1990). This benefit is attributed to there being a superior imaginal-enactive or multimodal representation for actions than for a verbal or phonological-based representation for oral repetitions. If this argument is true, given the involvement of working memory in following instructions, working memory should make different contributions to the two types of recall. This thus forms the second aim of this research.

In summary, the current research has two aims: first, to investigate the contribution of working memory to instruction-following task; and second, to confirm and also investigate the mechanism of the action advantage by observing its interplay with working memory components.

Dual task methodology and hypotheses

Dual task methodology is commonly used to separate the contributions of working memory components underlying the multicomponent working memory model (Baddeley & Hitch, 1974). The underlying logic is that tasks using the same cognitive components compete for resources, hence simultaneously processing the two tasks will lead to a decrement of performance; in contrast, tasks using different components will not (Baddeley, 1986).

A series of experiments were conducted using the dual-task methodology with the purpose of isolating subcomponents in the working memory. In order to understand the formation of the representation of instructions, all interference tasks disrupted the encoding and maintaining stage of instructions without impeding the recall.

To explore the phonological loop, the articulatory suppression task was chosen as it is a well established interference task that selectively impairs the phonological loop by preventing rehearsal (Baddeley, et al., 1984). This task involved participants repeating numbers continuously throughout the encoding stage of instruction. The phonological loop was expected to be highly involved in following spoken instructions, as it functions both as a passive storage of instructions in a phonological form and a rehearsal mechanism to refresh the fading memory trace. The visuospatial sketchpad may help track the spatial sequence of intended actions on associated objects. Given the complexity of the visuospatial sketchpad and its close relationship with the central executive, both eye-closure technique and spatial tapping task were used. As stated before, updating coming new information, processing sequential information, binding movements with objects, and monitoring one's own actions, are all elements requiring the successful functioning of the central executive. The central executive was therefore expected to be highly involved. The backward counting task requiring a continuous decrease of digits was used to disrupt the central executive additionally to the phonological loop.

Outline of the experiments

The research presented in this thesis began with investigating the involvement of working memory in spoken instructions using a computer-based task (Chapter 2). Participants listened to the instructions involving series of actions, e.g. 'click the flag drag the star onto the triangle click the arch drag the chevron onto the cross'. They were required to either use a mouse to drag and click the geometric shapes on the screen (action recall) or to repeat the instructional sentence back (verbal recall). Experiment 1 examined the roles of the phonological loop and central executive, and also compared two types of recall. Experiment 2 set the type of recall as a between-subject factor to prevent carryover or practice effect.

Chapter 3 addresses the possibility of action advantage in a rich task environment (Experiment 3). A task involving colourful objects and more variations of movements in a three dimensional world was used. Participants listened to instructional sentences such as ‘push the black pencil, and spin the green eraser, and touch the red pencil, and push the blue ruler, and touch the white eraser’, and were required to either recall the sentence or act upon the objects. The involvement of the phonological loop and the central executive was investigated using the articulatory suppression and backward counting tasks respectively.

Chapter 4 continues to explore the role of three working memory components in a rich environment using the 3D instructional task, with a focus on the visuospatial sketchpad. Experiment 4 used a simple spatial tapping task to disrupt the visuospatial sketchpad. Experiment 5 required participants to close their eyes when listening to the instructions, therefore blocking the encoding of visual and spatial information.

Chapter 5 aims to extend the findings of working memory in spoken instructions to written instructions. Action phrases like ‘push red box, pick up black pencil, put it into yellow bag, touch red pencil, spin blue ruler’ were presented on a computer screen in separate rows. Experiment 6 examines the role of the phonological loop using the articulatory suppression task and the contribution of the central executive using the backward counting task. Experiment 7 investigates the visuospatial sketchpad with a complex spatial tapping task. The final chapter brings together all the findings of the seven experiments, and discusses the limitations as well as the implications of this study.

Chapter 2

Following spoken instructions in a computer-based task

Introduction

The purpose of the two experiments in this chapter is to investigate the involvement of working memory components in following spoken instructions. Two main issues were addressed in these experiments. The first issue was what specific contributions of working memory components if any, contribute to following instructions, as shown in the literature (Allen, 2009; Allen & Gathercole, 2008; Engle, et al., 1991; Gathercole, et al., 2008; Kim, et al., 2008). The second issue concerns whether working memory mediates the phenomenon known as action advantage, which suggests that there is a substantial benefit to carrying out instructions in actions than simply repeating the instructions verbally (Allen, 2009; Allen & Gathercole, 2008; Gathercole, et al., 2008).

The issue of the precise contribution of components of working memory in following instructions was investigated using concurrent tasks known to selectively impair the subcomponents of working memory. Articulatory suppression is known to prevent rehearsal component of the phonological loop (Baddeley, et al., 1975). The backward counting task taps both the central executive and the phonological loop, and the decrement of the phonological loop was matched to show the specific contribution of the central executive (Allen, et al., 2006; Postma & De Haan, 1996). This was achieved by comparing the performance in the backward counting condition with the performance in the articulatory suppression condition. In each case, the concurrent tasks occurred during the presentation of the instructions, which was prior to the commencement of recall.

The second issue that recalling instructions by actions is better than oral repetition was tested by contrasting the accuracy of the two types of recall. My interest also lies in the extent to which the two concurrent tasks significantly influenced the accuracy with which participants could actually repeat or perform the action. These patterns of interference would provide important novel information on the extent to which subcomponents of working memory contributes to remembering instructions.

The research in this thesis started with spoken instructions as these are common in everyday life, especially in the situations of giving flexible instructions applicable to a wide range of people, like pre-reading children, elderly people, and clinical patients as well as typical adults. For example, teachers often give oral commands to guide children in classroom activities, such as 'put your sheet on the green table, put your pencil away and come and sit on the carpet'. Step-by-step instructions are also seen in typical learning, for example, in a maths class, where an instruction might be 'look at the two numbers. Take away the number at the bottom from the one at the top, and write down the answer under the line '(Gathercole & Alloway, 2008). There are also many instances of instructions in adult life, such as in driving lessons, which are often delivered by demonstration with additional oral explanations.

The version of the task used in the present experiments employed simple shapes such as circles and squares that were displayed simultaneously on the computer screen. Participants were required to follow spoken instructions to carry out a sequence of actions such as 'click the flag, drag the star onto the triangle'. They then either repeated the instructions or attempted to follow them by actions using the mouse to manipulate the shapes on the screen.

Experiment 1

Introduction

The main aim of the experiment was to investigate the role played by the two components of working memory, the phonological loop and the central executive in following instructions.

Several studies have provided evidence for the involvement of the phonological loop in following spoken instructions, such as the significant correlation between the ability to follow instructions by actions and the verbal rehearsal task (digit recall) in children (Gathercole, et al., 2008), and also the direct evidence using the dual task methodology, i.e. a significant decrement of recall was observed when the phonological loop was interfered by the articulatory suppression task (Allen, 2009; Allen & Gathercole, 2008). The phonological loop may contribute to an individual's ability to follow instructions in several ways. First, there is evidence showing that auditory sound gains access to the store automatically (Hanley & Broadbent, 1987; Neath, et al., 1998); hence instructions presented in auditory format may enter the phonological store easily. Second, because the phonological store tends to decay rapidly as time passes, and instructional sentences have to take a period time before recall, hence they would not be easily maintained in the store. Therefore, in order to maintain the representations of the lengthy sentences within the phonological loop, rehearsal would be required to offset the rapid decay within the phonological store. Articulatory suppression is known to disrupt the subvocal rehearsal (Baddeley, 1975), and on this basis, it is predicted that the recall would be impaired in the articulatory suppression condition.

The articulatory suppression task is also the interfering task which proved to be effective in Allen's study (2009), in which participants listened to the spoken instructions containing sequences of actions on laminated cards depicting simple geometric shapes, and then recalled the instructions either by repeating the sentence or manipulating the cards by hands. In the articulatory suppression condition, participants repeated the three-digit numbers continuously during the presentation of instructions. The articulatory suppression was found to disrupt both types of recall significantly, suggesting that the phonological loop is underlying the encoding of spoken instructions regardless of the subsequent type of recall. To make the findings more comparable to Allen's findings, the same concurrent task, that is the articulatory suppression task, was used in this research.

The central executive may play a number of different roles when a sequence of spoken instructions is being remembered. This includes paying selective attention to the intended objects, forming a mental representation that links specific movements to target objects, and keeping track of what has been done and what has yet to be done. The involvement of the central executive is supported by correlation studies requiring actions upon stationery objects (Gathercole, et al., 2008) and also Allen's task in which recall was impaired by the demand of counting the three-digit number backward (Allen, 2009). The backward counting task was selected to specifically impair the central executive. Both the backward counting task and the articulatory suppression task involve the spoken production of sequences of numbers. In addition, the backward counting task requires accessing knowledge of number, applying subtraction rules, and retaining the most recently generated number. Therefore, the contribution of the phonological loop has to be partialled out, which was achieved by comparing the performance in the backward counting task to the articulatory suppression task. The difference between the two corresponds to the specific role of the central executive.

Therefore, three types of conditions were formed: the articulatory suppression condition, the backward counting condition, and also a baseline condition which served as a comparison condition for the articulatory suppression condition. In order to compare the effect of two concurrent tasks on recall, scores representing the articulatory suppression effect and backward counting effect were calculated. The articulatory suppression effect was the difference between the accuracy of recall in the articulatory suppression condition and the baseline condition, representing the deduction on recall by suppressing the rehearsal. The backward counting effect was the difference between the accuracy of recall in the backward counting condition and the articulatory suppression condition, representing the specific disruption imposed by the demand of central processing after excluding the retaining function of the phonological loop.

The second aim of this experiment was to establish that the action advantage previously reported in the two instructional tasks would also be present in this paradigm. If it was, investigating the extent to which it is disrupted by either articulatory suppression or backward counting would be another interest of this study. Therefore, people were asked to recall either by actions or by repetition. The differential disruption by articulatory suppression or backward counting, if any, would be shown as the interactions between the concurrent tasks and the types of recall.

In summary, three hypotheses were tested in this experiment. First is that the phonological loop supports the maintenance of spoken instructions. If this is the case, the articulatory suppression task should lead to impairment in recall. The second hypothesis is that the central executive is involved in encoding spoken commands by paying selective attention to intended objects, linking movement with target objects, and keeping track of the status quo of online objects. According to this hypothesis, the backward counting task will disrupt the recall, which is reflected by the inferior performance comparing to the articulatory suppression conditions, serving as a control for rehearsal in the backward counting task. The third hypothesis is that the recall by actions will be superior to the simple repetition, that is, the advantage of action recall compared to the verbal recall. The interaction between working memory and recall is still in its exploration, for this reason, no specific hypothesis is made in the first experiment.

Method

Participants

Twenty-four native English speakers at the University of York were recruited through an electronic booking system, and they took part in the experiment in exchange for course credit or an honorarium of £6. There were 21 females and 3 males, aged from 18 to 32, with a mean age of 20.17.

Materials

Each instruction involved actions on six geometric shapes drawn from a sample pool containing eight types of basic geometric shapes, i.e. circle, diamond, star, cross, triangle, chevron, arch, and flag. These were single-line standard shapes that were the same as the stimulus materials in Allen's experiment (2009). Giving the limited range of mouse-based actions, only two types of movements were included, i.e. 'click' and 'drag…onto…'.

An instruction was created as a series of action phrases without using any conjunction word. For example, a typical instruction was, 'click the flag drag the star onto the triangle click the arch drag the chevron onto the cross'. This type of instruction was similar to imperative sentences, which assume a first-person perspective. Each instruction contained 18 words, and a repetition of the same shape was not allowed within an instruction. Each instruction contained six actions and six steps. Although the ‘drag…onto…’ action is likely to be perceived and performed as one action, separate errors for ‘drag’ and ‘onto’ can occur, such as dragging the correct shapes onto the wrong target shapes; ‘drag…onto…’ was therefore counted as two separate actions in this study.

The sounds of the shape and action words were recorded by a native English speaker in a flat tone, and were stored as individual sound files. During the presentation of an instruction, the sound file of each word was evoked according to a prewritten stimuli list. Each word lasted 500 ms with an average 340 ms gap between the words, and the duration of an instruction was around 15 seconds.

The stimuli of the visual array were presented on a 15-inch screen of an Apple laptop computer, with eight shapes aligned in two rows and four columns (see Figure 2.1). These shapes could be dragged and clicked using a mouse. The effects of these actions were presented as animations of a picture of a small hand representing the mouse on the screen. For example, the action ‘click the circle’ required participants to move the mouse until the small hand on the screen was on the circle; the selection key of the mouse was then pressed, and a black square around the circle would show to indicate that the circle had been clicked. The drag action, for instructions such as ‘drag the cross onto the star’, required participants to first click the cross, then move the mouse towards the star without releasing the selection key until the cross had reached the star, and finally release the key press. This was indicated by the cross now totally covering the star, representing the completion of the ‘drag…onto’ action.

An instructional list containing fourteen trials was constructed (see Appendix 1), with two practice trials, ten formal trials, and two spare trials for unexpected interruption. This list was used across conditions. The three-digit numbers for the articulatory suppression and backward counting conditions were generated randomly by the Supercard program during its running.

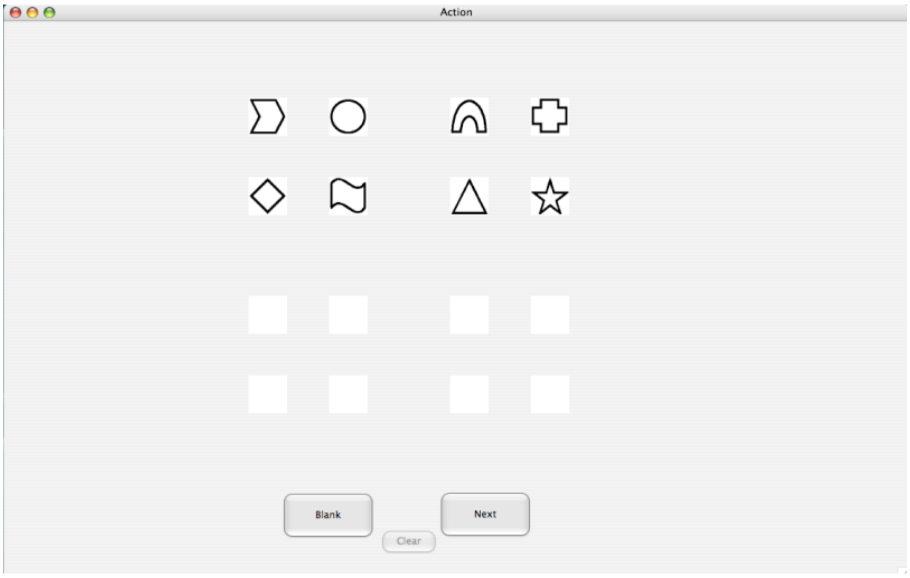

Figure 2.1 Visual display of the computer-based instruction task in Experiment 1.

A debriefing questionnaire was developed to investigate the subjective difficulty and the strategies employed in each condition. The difficulty concerning both the memory task for instructions and the secondary interference task were both rated using a 5-point Likert scale from very easy (point 1) to very difficult (point 5). Each condition contained ratings of difficulty in the encoding stage (e.g. remember the instructions) as well as in the retrieval stage (e.g. repeat or perform the instructions in orders). The articulatory suppression and the backward counting conditions required additional ratings on the interference task, i.e. ‘repeat the numbers when listening to the instructions’ and ‘count the numbers backwards when listening to the instructions’. Any strategies which the participants may have used to help them accomplish this were investigated by the question ‘if you are using any strategy, please state this’ (see Appendix 2).

Design

In a 3 $\times$ 2 within-subject design, one independent variable was concurrent task, containing three conditions: baseline, articulatory suppression and backward counting. The other independent variable was type of recall, i.e. verbal and action recall. The main dependent variable was the mean number of correct actions per instruction sequence. Other measurements included elements (shapes and movements), proportion of order errors, and also the percentage of correct trials in each serial position.

Procedure

The experiment was carried out in a quiet room. Upon arrival, participants were introduced to the task and asked to sign a consent form, and familiarized with the names of the geometric shapes. All participants finished six conditions. The verbal and action recall conditions were counterbalanced in sequence.

In each condition, the participants first read the task requirement, which emphasised the importance of recall in correct serial order. Participants first finished two practice trials, and if they failed at these, they were given another practice trial. Each participant finished six conditions.

In the baseline condition, participants first saw a bar containing the words 'press space bar to continue' on the computer screen. Two seconds after the space bar was pressed, the visual display of the shapes showed on the screen. 500 ms after the appearance of the visual display, the spoken instruction began to play. Participants listened to the instruction while looking at the array of the geometric shapes on the screen. Participants were not allowed to perform any actions during this stage. At the end of the instruction presentation, a blank screen was shown for one second, followed by a beep sound indicating the beginning of recall. In the verbal recall conditions, participants repeated the instructional sentence, whereas in the action recall conditions, they used the mouse to click and drag on the shapes. When the participants finished recall, they pressed the button 'next' at the bottom of the screen, which triggered the next trial.

In the articulatory suppression condition, participants first saw a three-digit number on the computer screen, and were required to repeat the numbers aloud at a rate of two seconds per cycle. The digits lasted two seconds on the screen, followed by a one-second blank screen, then the appearance of an array of shapes. After 500 ms, the spoken instructions began. Participants were instructed to continue repeating the numbers at the paced rate during the delivery of the spoken instruction, and a one-second blank screen followed until the beep sound. After the beep sound, participants stopped repeating or back-counting the numbers and began the recall. The procedure for the backward-counting condition was similar to the articulatory suppression condition, except that participants decreased the three-digit number by three continuously, for example, 3-5-8, 3-5-5, 3-5-2, until the beep sound.

The visual display was provided throughout the encoding and the recall stage. These shapes remained in the same locations within a single trial, but were varied randomly between trials to ensure novelty. To prevent mistakenly filling the location of a forgotten object with a later object, participants were required to indicate the forgotten action; participants clicked a blank tab at the bottom of screen in the action recall conditions, whereas in the verbal recall conditions they simply said the word 'blank'.

The content of recall was noted as correct or wrong by the experimenter in the verbal recall conditions, whereas the movements of clicking and dragging were recorded automatically by the Supercard software in the action recall conditions. Once recorded, the experimenter signalled the participant to start the next trial. Participants took a short break between conditions and were given the questionnaire at the end of the experiment.

Results

Actions

The serial recall of actions was scored by averaging the correct actions per instruction across the formal ten trials. An action is defined as a ‘chunk’ of elements containing items in the environment and movements carried out on them. An action was scored as correct only when both the combination of movement and shape, and its serial position, were correct. Given that there were six actions in one instruction, the maximum score was six. The means and the standard deviations of actions as functions of concurrent tasks and type of recall are displayed in Table 2.1.

Table 2.1 Means (and standard deviations) of actions in Experiment 1

	Verbal recall	Action recall	Means
Baseline	4.44 (1.09)	4.30 (0.78)	4.37 (0.83)
Articulatory suppression	3.58 (1.13)	4.00 (1.16)	3.78 (1.01)
Backward counting	1.30 (0.78)	1.21 (0.79)	1.25 (0.71)
Means	3.11 (0.78)	3.17 (0.76)	3.14 (0.74)

A 3 × 2 (Concurrent task × Recall type) ANOVA showed significant main effect of concurrent task, F (2, 23) = 222.707, $p < 0.001$, $\eta_p^2 = 0.906$, MSE = 0.592, but there was no significant difference between verbal recall and action recall, F (1, 23) = 0.454, $p = 0.507$, $\eta_p^2 = 0.019$, MSE = 0.326. The interaction between concurrent task and recall type was approaching significance, F (2, 46) = 2.744, $p = 0.075$, $\eta_p^2 = 0.107$, MSE = 0.423.

A planned contrast indicated evident articulatory suppression effect, F (1, 23) = 17.711, $p < 0.001$, $\eta_p^2 = 0.461$, MSE = 0.413, and evident backward counting effect, F (1, 23) = 199.808, $p < 0.001$, $\eta_p^2 = 0.897$, MSE = 0.772. Individual scores of the articulatory suppression and backward counting effect were calculated for each participant. The comparison of the two effects and their interactions with recall type was examined with a 2 × 2 (Effect × Recall type) ANOVA. The backward counting effect was significantly larger than the articulatory suppression effect, F (1, 23) = 51.441, $p < 0.001$, $\eta_p^2 = 0.691$, MSE = 1.779, but there was no significant effect of recall type, F (1, 23) = 0.044, $p = 0.836$, $\eta_p^2 = 0.002$, MSE = 0.276. There was a marginally significant interaction between the effect and recall type, with a greater articulatory suppression effect in the verbal recall and a greater backward counting effect in the action recall, F (1, 23) = 4.052, $p = 0.056$, $\eta_p^2 = 0.150$, MSE = 1.711.

One-tailed dependent *t*-tests with Bonferroni corrections found significant articulatory suppression effect in the verbal recall conditions, t (23) = 4.412, $p < 0.01$, but not in action recall conditions, t (23) = 1.486, $p = 0.304$. The backward counting effects were significant in both types of recall conditions (*ps* < 0.01).

Elements

Each action contains two elements, movement and shape. Accuracy for movement and shape was calculated independently, with each scored as correct if recalled in the appropriate serial position. For example, if the third action in an instruction was 'click the arch', and the participant recalled this as 'click the circle', the recall of movement was considered correct but the recall of the shape was considered incorrect. The score of movement and shape were calculated by averaging the number of correct ones in each instruction; these ranged from 0 to 6. However, the scores of movement and shape cannot be compared directly because they are at difference chance levels; for movement, the chance of guessing it correctly is 50 percent, whereas the chance level for shape is one in eight, i.e. 12.5 percent. The means and standard deviations as functions of concurrent tasks and type of recall are shown in Table 2.2.

Table 2.2 Means (and standard deviations) of elements in Experiment 1

		Verbal recall	Action recall	Means
Movement	Baseline	5.21 (0.78)	4.92 (0.59)	5.06 (0.60)
	Articulatory suppression	4.30 (1.03)	4.65 (1.05)	4.47 (0.94)
	Backward counting	2.37 (0.88)	2.30 (0.96)	2.33 (0.82)
	Means	3.96 (0.74)	3.96 (0.69)	3.96 (0.68)
Shape	Baseline	4.67 (0.93)	4.63 (0.76)	4.65 (0.77)
	Articulatory suppression	4.01 (1.05)	4.32 (0.97)	4.16 (0.84)
	Backward counting	1.70 (0.92)	1.69 (0.97)	1.69 (0.81)
	Means	3.46 (0.74)	3.54 (0.75)	3.50 (0.69)

Note. The chance levels for movement and shape are different.

A 3 × 2 × 2 (Concurrent task × Recall type × Element) ANOVA was conducted. There was significant main effect of concurrent task, F (2, 46) = 253.04, p < 0.001, η_p^2 = 0.917, MSE = 0.866, element, F (1, 23) = 50.396, p < 0.001, η_p^2 = 0.687, MSE = 0.294, but no significant main effect of recall type, F (1, 23) = 0.175, p = 0.680, η_p^2 = 0.008, MSE = 0.645. There was a significant interaction between element and concurrent task, F (2, 46) = 4.614, p = 0.015, η_p^2 = 0.167, MSE = 0.152, but there was no other two-way or three-way interactions (*ps* > 0.05). In specific, the backward counting was more disruptive to shape than to movement, F (1, 23) = 9.081, p = 0.006, η_p^2 = 0.283, MSE = 0.290, while the articulatory suppression effect, however, was similar for shape and movement, F (1, 23) = 0.960, p = 0.337, η_p^2 = 0.040, MSE = 0.269. There was no other two-way or three-way interaction (*ps* > 0.05).

One-tailed dependent *t*-tests with Bonferroni corrections found significant articulatory suppression effect in both movement and shape (*ps* < 0.01), and significant backward counting effect in both movement and shape (*ps* < 0.01).

Binding

Taking a different perspective, an accurate action can also be seen as a correct combination of the elements (shape and movement). For example, the action ‘click the circle’ was scored as correct only when both the movement ‘click’ and the shape ‘circle’ were correct. An element, however, was scored as correct without consideration of the other element in the same action. Taking the same example, if the participant recalled ‘click the star’, the movement was scored as correct whereas the shape was scored as incorrect. Therefore, recalling an action accurately was more difficult than recalling an element correctly. The effort for binding two elements in an action can be reflected as the difference between the scores for an action and its elements. For instance, the difference in scores for actions and shapes reflects the cost of binding the correct movements to the corresponding shapes.

It is therefore theoretically interesting to test whether the central executive contributes to the binding (Baddeley, 2000; Baddeley, et al., 2011). Studies have showed controversial results regarding the role of the central executive in binding (Allen, et al., 2006; Brown & Brockmole, 2010). If the central executive indeed helps the binding of movement and shape in an action, then the backward counting should be more disruptive to memory for actions than for elements, reflected as an interaction between backward counting and binding cost. Moreover, if the role of binding played by the central executive differs in verbal and action recall, it should be shown as a three-way interaction.

It is worth noting that the element was part of the bound entity, and hence their scores linked probabilistically; that is, the scores for colours and shapes were not independent of scores for an action. Nevertheless, the test can still offer insights into the role of the central executive in binding.

The binding of a movement to an object was tested by a 2 × 2 × 2 (Backward counting × Binding × Recall type) ANOVA. The variable backward counting included the articulatory suppression and backward counting conditions, the variable binding included action and object, and the variable recall contained verbal and action recall. Results showed significant main effect of backward counting, F (1, 23) = 263.229, $p <$ 0.001, $\eta_p{}^2$ = 0.920, MSE = 1.143, binding, F (1, 23) = 58.282, $p < 0.001$, $\eta_p{}^2$ = 0.717, MSE = 0.136, but no significant effect of recall type, F (1, 23) = 1.709, $p = 0.204$, $\eta_p{}^2$ = 0.069, MSE = 0.670.

There was no significant interaction between backward counting and binding, backward counting effect was similar in action and object, that is, binding movement to object did not require central executive, F (1, 23) = 0.299, $p = 0.590$, $\eta_p{}^2$ = 0.013, MSE = 0.163. There was no other two-way or three-way interaction ($ps > 0.05$) .

Serial positions

Each position was coded for the percentage of correct trials, ranged from 0 to 1. The serial position curves as functions of concurrent tasks and type of recall are shown in Figure 2.2.

A 3 × 2 × 6 ANOVA (Concurrent task × Recall type × Serial position) was conducted, and there was significant main effect of concurrent task, F (2, 46) = 223.870, $p < 0.001$, $\eta_p{}^2$ = 0.907, MSE = 9.899, serial position, F (2.74, 62.91) = 27.350, $p < 0.001$, $\eta_p{}^2$ = 0.543, MSE = 3.819, but no significant main effect of recall type, F (1, 23) = 0.371, $p = 0.548$, $\eta_p{}^2$ = 0.016, MSE = 5.498. The interaction between concurrent task and recall type was approaching significance, F (2, 46) = 2.702, p = 0.078, $\eta_p{}^2$ = 0.105, MSE = 7.149. Position interacted with concurrent task, F (3.33, 76.62) = 6.297, $p < 0.001$, $\eta_p{}^2$ = 0.215, MSE = 6.609, but not with recall type, F (3.24, 68.32) = 0.233, p = 0.298, η_p^2 = 0.051, MSE = 2.623. There was a significant three-way interaction, F (4.48, 102.98) = 5.901, p < 0.001, η_p^2 = 0.204, MSE = 3.618.

The post-hoc tests with Bonferroni corrections indicated that the decrease was significant between positions 1 and 2, and between positions 2 and 3 (ps < 0.01), but not between other adjacent positions (ps > 0.05). Planned contrasts showed significant interaction between articulatory suppression and position 3 - 4 (p = 0.002) and position 4 - 5 (p = 0.039). Backward counting effect interacted with position 3 - 4 (p = 0.001) and position 5 - 6 (p = 0.004).

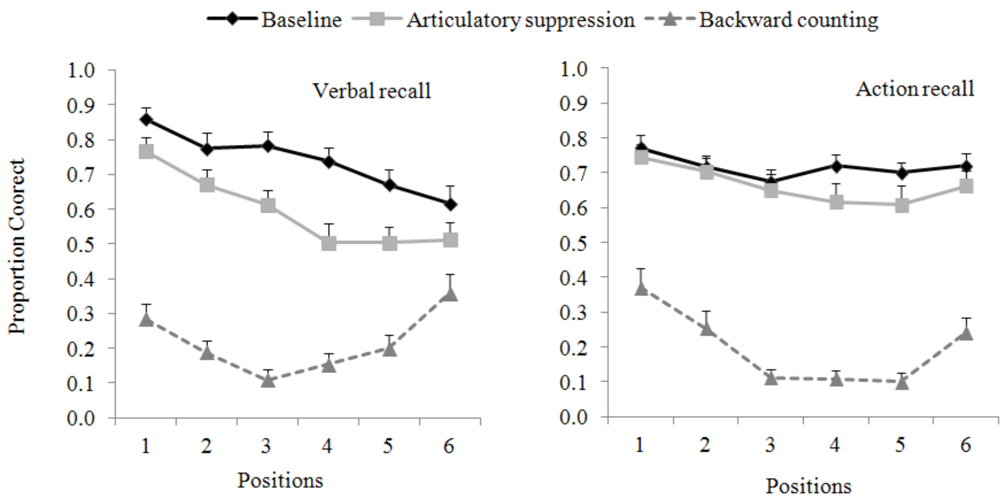

Figure 2.2 Serial position curves (means and standard errors) as functions of concurrent tasks and type of recall in Experiment 1.

Practice effect

One possibility for the absence of the action advantage may be attributed to the carryover effect. When participants engaged in an action recall then also engaged in a verbal recall, they may have became familiar with the way representing actions for enactment. As this type of mental representation was found to be effective, similar type of mental representation was generated for the oral repetition.

The same condition that tested first and that tested last in a sequence was compared; for example, four participants finished the baseline-verbal recall condition first and the other four participants finished this condition as the last condition. The means and standard deviations of actions as functions of concurrent tasks, type of recall and sequence of conditions are presented in Table 2.3.

Table 2.3 Means (and standard deviations) of actions in go-first conditions and go-last conditions in Experiment 1

	Recall	Go as first	Go as last
Baseline	Verbal	3.20 (1.58)	5.10 (0.42)
	Action	4.07 (0.67)	4.05 (0.88)
Articulatory suppression	Verbal	2.10 (1.36)	4.17 (0.69)
	Action	3.81 (0.77)	4.65 (0.45)
Backward Counting	Verbal	0.98 (0.70)	1.53 (0.31)
	Action	1.26 (0.96)	1.30 (0.70)

Figure 2.3 presents the performance of recalls in six conditions. The differences between the black and white bars stand for the gains from practice. It should be noted that the practice gains were the average of the participants rather than the individual gain, because each participant finished one condition only once. One striking difference in the graph was the greater gains in verbal recall conditions compared with the action recall conditions.

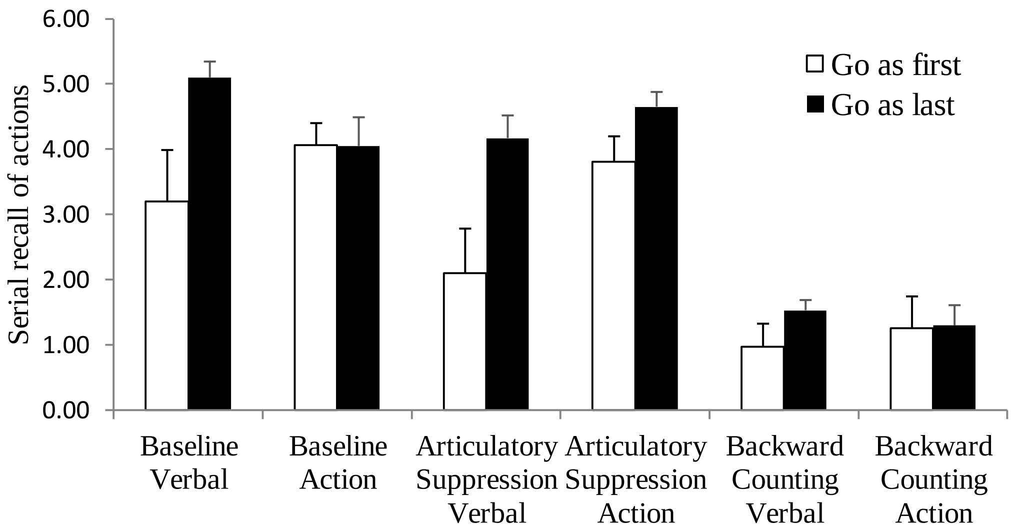

Figure 2.3 Action (with standard errors) as functions of concurrent tasks, recall type, and sequence of conditions in Experiment 1

Strategy report

Eighteen of the twenty-four participants indicated that they had intentionally implemented strategies. Among the eighteen respondents, eleven indicated using multiple strategies. There were seven different strategies in total, and the most common strategy used was mentally drawing linking lines between the shapes as their names were mentioned in the sequence. A similar strategy was the imagining of the self clicking and dragging shapes when listening to the instructions. Grouping the actions and focusing on the beginning and the end of an action sequence were both strategies related to organizing and optimizing the encoding process. In the interference conditions, for example, the articulatory suppression condition, there were occasions on which participants tried to decrease the interference by thinking less about the suppression task. One participant created her own coding system to help recall and used acronyms to stand for shapes and numbers for movements.

The numbers and percentages of the responders using a specific strategy are summarized in Table 2.4. Because some participants indicated using multiple strategies, the sum of the percentages exceeds 100%.

Table 2.4 Self-report strategies in Experiment 1

Strategies	Count	Percentage
Mentally draw lines between objects	11	61%
Verbal rehearsal	4	22%
Imagining carrying out the action	3	17%
Decreasing interference	3	17%
Group actions	2	11%
Focus on beginning and end of sequence	2	11%
Use acronyms	1	6%

Ratings of difficulty

In each condition, participants rated the difficulty felt in the encoding stage and the retrieval stage of the task using a 5-point Likert scale. The means and standard deviations of rated difficulty as functions of concurrent tasks and type of recall are presented in Table 2.5.

Table 2.5 Means (and standard deviations) of difficulty ratings in Experiment 1

	Recall	Encoding	Retrieval	Interference task
Baseline	Verbal	2.88 (0.99)	2.96 (1.08)	NA
	Action	3.04 (1.00)	3.13 (1.19)	NA
Articulatory suppression	Verbal	3.71 (0.96)	3.79 (0.98)	2.04 (1.27)
	Action	3.58 (0.83)	3.71 (0.81)	2.04 (1.23)
Backward counting	Verbal	4.88 (0.34)	4.71 (0.55)	4.17 (0.96)
	Action	4.79 (0.51)	4.75 (0.44)	4.46 (0.78)

Note. NA stands for non-applicable, because there was no interference task in the baseline conditions

The 3 × 2 × 2 (Concurrent task × Recall type × Stage) ANOVA found significant main effect of concurrent task, F (1.52, 35.04) = 86.083, $p < 0.001$, $\eta_p^2 = 0.789$, MSE = 1.179, but no significant effect of recall type, F (1, 23) = 0.023, $p = 0.882$, $\eta_p^2 = 0.001$, MSE = 0.615, or stage, F (1, 23) = 0.115, $p = 0.738$, $\eta_p^2 = 0.005$, MSE = 0.483. Planned contrast found significant articulatory suppression effect, F (1, 23) = 43.038, $p < 0.001$, $\eta_p^2 = 0.652$, MSE = 0.272, and significant backward counting effect, F (1, 23) = 73.340, $p < 0.001$, $\eta_p^2 = 0.761$, MSE = 0.384. There was no two-way or three-way interaction (*ps* > 0.05).

Discussion

This experiment investigated the role of working memory in following spoken instructions in a task involving remembering instructions on clicking and dragging simple shapes on a computer screen. The two concurrent tasks used were the articulatory suppression task and the backward counting task. The results showed that both articulatory suppression and backward counting tasks impaired recall of actions, which were consistent with previous findings (Allen, 2009).

The disruptive effect of articulatory suppression suggests that the phonological loop is involved in following spoken instructions. Interestingly, there was a trend of a larger articulatory suppression effect in verbal recall compared to action recall, implying a greater need of the phonological loop for constructing a verbal-based representation when a verbal-based repetition was required. Specific, the articulatory suppression was present in the verbal recall condition but absent in the action recall condition, indicating that the phonological loop contributes when an oral repetition is required, but is less likely to play a part when an enactment is needed. The difference can also been seen from the shapes of the serial position curves (see Figure 2.2). In the verbal recall condition (left panel), for each position, the accuracy of recall in the articulatory suppression condition was lower than that in the baseline condition, suggesting the involvement of the phonological loop was throughout the sequence. In contrast, in the action recall condition (right panel), the articulatory suppression effect was only evident in later positions, implying that earlier actions were remembered without the assistance of the phonological loop, probably via some other more efficient means (perhaps the visuospatial storage). Further rehearsal was needed when the other type of storage was insufficient; the phonological loop was thus recruited again to help memorize the actions later in a sequence.

The larger disruptive effect of the backward counting task than the articulatory suppression task also indicates the involvement of the central executive. One hypothesized role of the central executive was that it links a specific movement to the intended object. However, in this experiment, counting backward similarly disrupted memory of action and memory of shape, implying that binding movements to shapes did not demand the central executive. The effect of backward counting can also be spotted from the shape of the serial position curves, which differed from those in the baseline and articulatory suppression conditions (see Figure 2.2). The bow-shaped curves with a large recency effect implied that participants were strategically abandoning the intermediate actions in the sequence, in order to save some time for later actions before they were forgotten.

Moreover, even after controlling the contribution of the phonological loop, the disruptive effect of the backward counting task was still larger than the articulatory suppression effect, suggesting a greater contribution from the central executive than from the phonological loop. This result also suggests that following instructions is a task requiring higher cognitive functions than simple maintenance of information. Subjective ratings of difficulty generated at the end of the experimental session also corroborated the finding of the objective measures. Specifically, the articulatory suppression condition was rated as more difficult than the baseline condition, and the backward counting condition was rated as more difficult than the articulatory suppression condition, suggesting the involvement of the phonological loop and central executive.

Some may argue, however, from the attentional view of working memory (Cowan, 1999), the difficulties experienced by participants reflect the attentional demand in the two concurrent tasks; therefore, their disruptive effects may also indicate that different attentional resources are left for the main task, i.e. remembering the instructions. The greater contribution of the central executive compared to the phonological loop therefore implies that there is larger attentional disruption from a more difficult task relative to a simpler task. Nevertheless, it should be noted that this view is not incompatible with the multicomponent working memory model, which postulates that the central executive is an attentional system that is in charge of focusing, dividing and switching attention (Baddeley, 1996, 2007). Accordingly, the additional contribution from the central executive compared to the phonological loop may therefore reflect the great importance of attention in remembering instructions.

Contrary to previous studies (Allen, 2009; Allen & Gathercole, 2008), no action advantage was obtained in this experiment. Action advantage is assumed to arise from the different representations formed during encoding: a superior imaginal-enactive or multimodal representation for action recall compared to an inferior verbal representation for verbal recall (Koriat, et al., 1990). However, in this experiment, the larger practice effect in verbal recall compared to action recall implies the existence of the carryover effect; that is, that the participants relied on a verbal coding in the verbal recall condition initially, but, after they had had experience of using a more efficient multimodal representation to guide the actions in the action recall conditions, the verbal coding strategy was replaced by the multimodal representation approach, resulting in the increased percentage of verbal recall. Another possibility is that verbal recall benefited more from practice than action recall. The practice effect and carryover effect was also observed in a previous study (Koriat, et al., 1990, Experiment 3).

If this is true, separating the verbal and action recall may lead to the formation of different representations during encoding, and inferior verbal representation would lead to a poorer performance in verbal recall. Experiment 2 was designed to test this hypothesis. In this experiment, action and verbal recall were tested under two concurrent task conditions, the baseline and articulatory suppression conditions. The crucial difference from the Experiment 1 was that a between-subject design was applied to the action and verbal recall conditions, thereby eliminating the opportunity for the carryover effects described above, which may have overshadowed an underlying advantage of action recall to the verbal recall.

Experiment 2

Introduction

Experiment 1 provided strong evidence for the involvement of the phonological loop and central executive in following spoken instructions. However, the expected action advantage of recalling actions over the simple repetition of the instructions found in previous studies (Allen, 2009; Allen & Gathercole, 2008; Gathercole, et al., 2008)was not obtained. It is speculated that the lack of action advantage may be due to the carryover effect resulting from a within-subject design in Experiment 1.Experiment 2 was therefore carried out to rule out these effects by setting the recall type as a between-subject factor. If the similar performances in verbal and action recall were mainly caused by the carryover effect, then separating the two types of recall should lead to the acquisition of the action advantage.

There was a trend towards a larger disruption of the articulatory suppression in the verbal recall compared to the action recall, implying a greater involvement of the phonological loop in forming a representation for the verbal recall than for the action recall. Therefore, the articulatory suppression condition was included in this experiment to test this hypothesis.

Method

Participants

Sixteen native English speakers at the University of York were recruited through the electronic booking system in exchange for course credit or an honorarium of £2. None had taken part in the previous experiment. There were 13 females and 3 males, aged from 19 to 25, with a mean age of 20.71.

Materials

The materials were the same as those used in Experiment 1.

Design

In a 2 × 2 mixed design, concurrent task was set as a within-subject variable, including baseline and articulatory suppression conditions, and recall type was set as a between-subject variable, including verbal and action recall. The dependent variables were same as those in Experiment 1.

Procedure

The participants were randomly assigned into either the verbal recall or the action recall group. Other procedures were the same as those in Experiment 1 except that each participant finished only two conditions, which were counterbalanced.

Results

Actions

The calculation of the action scores was same as that in Experiment 1. The means and standard deviations of actions as functions of articulatory suppression and recall type are shown in Table 2.6.

Table 2.6 Means (and standard deviations) of actions in Experiment 2

	Verbal recall	Action recall	Means
Baseline	4.04 (1.15)	4.22 (1.01)	4.13 (1.05)
Articulatory suppression	3.36 (1.00)	3.72 (1.08)	3.54 (1.02)
Means	3.70 (1.03)	3.97 (0.93)	3.84 (0.96)

A 2 × 2 (Articulatory suppression × Recall type) ANOVA revealed significant main effect of articulatory suppression, F (1, 14) = 8.837, p = 0.010, $\eta_p{}^2$ = 0.387, MSE = 0.318, but no significant main effect of recall type, F (1, 14) = 0.300, p = 0.592, $\eta_p{}^2$ = 0.021, MSE = 0.971, and no significant interaction between articulatory suppression and recall type, F (1, 14) = 0.300, p = 0.659, $\eta_p{}^2$ = 0.014, MSE = 0.318.

One-tailed dependent t-tests with Bonferroni corrections showed that the articulatory suppression effect was not in the verbal recall conditions, t (7) = 3.080, p = 0.018, but nor in the action recall conditions, t (7) = 1.517, p = 0.174.

Elements

The calculation of elements was the same as that in Experiment 1. The means and standard deviations of movements and shapes as functions of articulatory suppression and recall type are shown in Table 2.7.

Table 2.7 Means (and standard deviations) of elements in Experiment 2

		Verbal recall	Action recall	Means
Movement	Baseline	4.88 (0.72)	4.93 (0.85)	4.91 (0.76)
	Articulatory Suppression	4.37 (0.88)	4.25 (0.97)	4.31 (0.90)
	Means	4.63 (0.78)	4.59 (0.63)	4.61 (0.68)
Shape	Baseline	4.24 (1.09)	4.30 (1.14)	4.27 (1.08)
	Articulatory Suppression	3.43 (0.98)	4.13 (1.10)	3.78 (1.07)
	Means	3.83 (0.97)	4.22 (0.94)	4.02 (0.94)

A 2 × 2 × 2 (Articulatory suppression × Recall type × Element) ANOVA found significant main effect of articulatory suppression, F (1, 14) = 5.834, p = 0.030, $\eta_p{}^2$ = 0.294, MSE = 0.800, element, F (1, 14) = 31.782, p < 0.001, $\eta_p{}^2$ = 0.694, MSE = 0.172, but no significant main effect of recall type, F (1, 14) = 0.180, p = 0.678, $\eta_p{}^2$ = 0.013, MSE = 0.662. There was no two-way or three-way interaction (ps > 0.05).

One-tailed dependent *t*-tests with Bonferroni corrections found significant articulatory suppression effect in movement (p = 0.024) but not in shape (p= 0.080).

Serial positions

The scoring method of serial position was same as that in Experiment 1. The result of a 2 × 2 × 6 (Articulatory suppression × Recall type × Serial position) ANOVA showed significant main effect of concurrent task, F (1, 14) = 9.110, p = 0.009, $\eta_p^{\,2}$ = 0.394, MSE = 5.160, serial position, F (2.18, 30.54) = 6.216, p < 0.001, $\eta_p^{\,2}$ = 0.307, MSE = 49.896, but no significant effect of recall type, F (1, 14) = 0.295, p = 0.595, $\eta_p^{\,2}$ = 0.021, MSE = 2.718. There was no two-way or three-way interaction (*ps* > 0.05). Post hoc tests with Bonferroni corrections indicated that the decrement was only significant between positions 1 and 2 (p = 0.022), but was not significant between other adjacent positions (*ps* > 0.05).

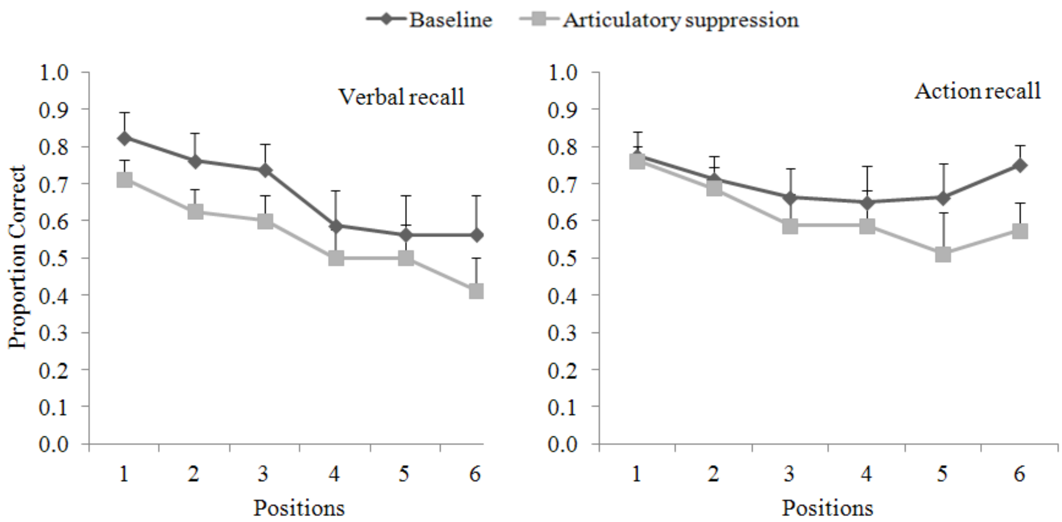

Figure 2.4 Serial position curves (means and standard errors) as functions of articulatory suppression and type of recall in Experiment 2.

Strategy report

Among the sixteen participants, eleven indicated intentionally implementing strategies in following instructions. There were a total of four different strategies used. As in Experiment 1, most participants tracked the objects and mentally drew the linking lines between the shapes mentioned in the sequence. The numbers and percentages of responders who indicated using specific strategies are summarized in Table 2.8.

Table 2.8 Self-report strategies in Experiment 2

Strategies	Count	Percentage
Mentally draw lines between objects	10	91%
Imagining carrying out the action	2	18%
Decreasing interference	1	9%
Group actions	1	9%

Ratings of difficulty

The means and standard deviations of ratings of difficulty in the encoding and retrieval stage as the functions of articulatory suppression and recall type are shown in Table 2.9.

Table 2.9 Means (standard deviations) of difficulty ratings in Experiment 2

	Recall	Encoding	Retrieval	Interference task
Baseline	Verbal	3.75 (1.04)	3.63 (0.74)	NA
	Action	3.38 (0.74)	3.50 (1.20)	NA
Articulatory suppression	Verbal	4.25 (0.46)	4.00 (0.93)	2.86 (0.17)
	Action	3.88 (1.13)	3.63 (0.98)	2.37 (0.92)

Note. NA stands for non-applicable, and there was no interference task in the baseline conditions

A 2 × 2 × 2 (Articulatory suppression × Recall type × Stage) ANOVA showed there was no significant main effect of articulatory suppression, F (1, 14) = 2.220, p = 0.158, $\eta_p{}^2$= 0.137, MSE = 1.013, no significant main effect of recall type, F (1, 14) = 0.914, p = 0.355, $\eta_p{}^2$ = 0.061, MSE = 0.427, and no significant main effect of stage, F (1, 14) = 0.427, p = 0.524, $\eta_p{}^2$ = 0.030, MSE = 0.585. There was no significant two-way or three-way interaction (ps > 0.05).

Discussion

This experiment failed to obtain the action advantage when the verbal and action recall was separated into two groups, excluding the cause of the carryover effect. To understand the lack of action advantage, it is worth comparing the task design in the two experiments to that in the previous studies which did obtain this difference.

Allen's laminated cards instructional task (2009) is considered first. This required a number of actions upon the laminated cards of geometric shapes, and the main difference from the current task was its involvement of a wider range of physically distinctive actions. These actions, including 'push', 'spin', 'touch', and 'drag', were ones that required direct contact with the objects, whereas the actions in the current task were indirect operations using a mouse to simulate movements of shapes on a computer screen. It is possible that the variation of actions was a crucial factor for the existence of action advantage. This is because compared to describing the movements with words, directly acting them out by hand is more intuitive and relatively easy. Therefore, an increased number of actions may cause more difficulty in orally describing these movements than directly performing them, leading to the action advantage.

The classroom instruction task (Gathercole, et al., 2008) required actions upon colourful stationery objects in a three-dimensional world, in contrast to the simple abstract line drawings of shapes that matched for size in the current task. These objects contained more semantic information, and their affordances may help prepare the actions. For example, it has been found that people tend to pre-shape their hand gestures before reaching for an object (Jeannerod, 1997e). The direct contact with the objects may also provide additional proprioceptive feedback that facilitates the speed and accuracy of future actions. Moreover, in the classroom instructional task, the objects were displayed in a three-dimensional world, which is more spatially distinctive than the two-dimensional surface of a computer screen in the current task.

Taken together, these comparisons indicate that the computer-based instruction task contained fewer variations in the types of actions and relatively abstract objects compared with the other two tasks that secured action advantage. Moreover, the current task required actions indirectly through a computer device rather than direct actions upon the objects in a three-dimensional world. It is assumed that the action advantage results from a superior multimodal representation for action recall compared to an inferior verbal-coded representation for verbal recall. This multi-modal encoding in the action recall might be weakened by the simplicity of the setting in this study. Therefore, perhaps, a rich cue environment is essential for the rise of action advantage. In other words, if action advantage is benefited mainly through forming a multi-modal encoding representation, an environment with rich cues would benefit action recall more than verbal recall.

In Experiment 1, there was a trend towards the verbal recall relying more on the phonological loop compared to the action recall. However, this trend was not replicated in the present experiment as there was no significant interaction between articulatory suppression and recall. Therefore, the postulation of a more verbal-based representation for repetition than for action is not supported in this computer-based instruction task. Despite the lack of interaction, the pattern of serial position curves (see Figure 2.4) still resembles that in Experiment 1. When oral repetition is required, sloping lines indicating a clear articulatory suppression effect can be observed in all serial positions (left panel); when recall by enactment was required, however, the position curves are relatively flat, with little suppression effect in the earlier positions, which increases gradually (right panel). Again, these results suggest that rehearsal is used differently in verbal and action recall.

Last but not east, it should be noted that there were only sixteen participants, which may decrease the power of detecting the effects. This may contribute to the null results, such as the absence of the action advantage. Nevertheless, the effect sizes of these null results were quite small; adding more participants is therefore unlikely to increase the chance of obtaining the significant results.

General Discussion

Summary of results

The first two experiments set out to test the involvement of working memory components in following spoken instructions. A computer-based instruction following task involving manipulation of geometric shapes was employed. Participants listened to the instruction and either repeated it back or used the mouse to act upon the shapes.

Taken together, Experiments 1 and 2 have established that both the phonological loop and the central executive contributed to participants' abilities to follow spoken instructions, with a greater contribution from the central executive. These results were consistent with previous research (Allen, 2009; Gathercole, et al., 2008). However, both experiments failed to obtain the action advantage, the recall of instructions was better by actions than by repetition. There was no interaction between working memory components and the type of recall.

Contributions of phonological loop and central executive

The contributions of the phonological loop and the central executive were reflected in the performance of both recall of actions and the recall of the individual elements. The similar articulatory suppression effect in movement and shape also suggests that the phonological loop supports the maintenance of different types of elements in an action in a similar way.

The greater involvement of the central executive compared to the phonological loop suggests the complexity of the cognitive process of following instructions. Memorizing a series of oral commands is more than simply retaining verbal materials; it requires the participation of higher cognitive functions. The unique role of the central executive probably relates to the attentional allocation, with greater attention given to encoding the difficult element in an action, i.e. shape.

Action versus verbal recall

Whereas previous studies showed the superiority of recalling instructions by actions compared to the verbal repetition of the instructions (Allen, 2009; Gathercole, et al., 2008), these experiments failed to obtain a difference between the two types of recall. The hypothesis of the action advantage is built on the assumption of a superior multimodal representation for actions compared to a verbal coded representation for spoken recall. The lack of action advantage in Experiment 1 was explained by the carryover effect, that is, that the participants might have found the encoding strategy for action recall useful and hence employed the same strategy for spoken recall of the instructions. To eliminate this carryover effect, the two types of recall were set as a between-subject factor in Experiment 2. If the lack of action advantage was indeed caused by the carryover effect, then preventing the transference of strategy should have led to the formation of different representations, a superior one for actions compared to an inferior one for repetition. However, the results of Experiment 2 failed to obtain the action advantage, and therefore did not support this hypothesis.

Comparisons of the task design in the current study with previous research implied the lack of action advantage might be due to the simplicity of the task setting.

Hypothetically, therefore, a task containing rich cues might enlarge the difference between the two types of representations in the verbal and action recall. It is assumed that a multimodal representation allows the integration of various dimensions of codes and simultaneous access to these codes, whereas a verbal representation suffers from representing multiple dimensions of elements in a sequential way (Koriat, et al., 1990). For example, frequent switches between the dimensions can be costly; ‘click the circle, click the triangle’ requires three switches between dimensions, from ‘movement’ to ‘object’ then again, ‘movement’ and ‘object’. Adding one dimension, such as colour, will increase the number of switches between dimensions’ for example, ‘click the red circle, click the blue triangle’ requires five switches. All in all, it is expected that action advantage should arise in a rich environment.

Subjective ratings of difficulty

In Experiment 1, the subjective ratings of difficulty reached the same conclusions as the objective measurements, with significant involvement of working memory components and no difference between verbal and action recall. Although the subjective ratings in Experiment 2 failed to show a significant articulatory suppression effect, a trend was observed from the descriptive data. In both experiments, there was no discernible difference in the ratings of perceived difficulty between the encoding stage and retrieval stage. This is not surprising, as the task might be perceived as a whole by the participants, or it could be that a weaker representation formed during encoding also caused more difficulty during the retrieval stage.

Although the subjective ratings showed the ability to reflect the similar findings of the subjective measurements, they did not provide additional information about the cognitive process of following instructions. Compared to objective measurements, subjective ratings tend to vary depending on individual criteria of difficulty; for this reason, the questionnaire of the subjective ratings of difficulty was not included in future experiments.

Strategy

In both experiments, visualizing the pattern by mentally drawing links of shapes in sequence was the most commonly used strategy. This is consistent with literature reporting that listeners tend to look at objects as they recognize the spoken words associated with them (Griffin, 2004). The spatial locations of objects might also have served as deictic pointers; that is, participants remember the locations of objects rather than objects themselves, and look back at the locations to retrieve detailed information of the objects (Spivey & Geng, 2001). This active control of eye movements might be directed by the central executive (Postle, et al., 2006), and the map of locations was probably retained by the visuospatial sketchpad.

Consistent with the lack of difference between verbal and action recall based on their performance, most participants indicated using similar strategies in the verbal and action recall conditions. However, it should be noted that the strategy report was based on the free report and some participants may have failed to notice, or forgotten, the strategies they actually employed in the task. Therefore, common mnemonic strategies should be provided as choices in the future questionnaire of strategies.

The next step

Among the many interesting possibilities raised by the first two experiments, the most urgent question was the discrepancy with previous research, that is, the lack of action advantage. A reasonable next step was to explore the possibility of action advantage being present in a rich environment. There were various ways of creating a more profuse task environment which could resemble the task settings in previous research.

In order to secure the acquisition of the action advantage, rather than changing potential contributing factors to the action advantage one at a time, it was decided that a task including all these potential contributors should be included, leading to a new paradigm of instructional task.

One of the purposes of these experiments was to explore the possible role of working memory in action advantage in an instruction-following task. The absence of this advantage in the two previous experiments made it impossible to investigate this. The next set of experiments therefore adopted an instruction-following paradigm in which a robust action advantage has already been established. This paradigm would involve actions in an environment with rich cues, including varieties of actions upon common objects in a three-dimensional world. The details of the design of this new paradigm will be introduced in the next chapter.

Chapter 3

Following spoken instructions in a rich environment

Introduction

The previous two experiments investigated the role of working memory in following instructions using a computer-based instructional task that involved clicking and dragging abstract shapes using a mouse device. These experiments established that there is a substantial involvement of working memory in this, with a moderate contribution from the phonological loop and a greater contribution from the central executive. However, the action advantage, the superiority of recall by actions to recall through oral repetition, was not obtained. Therefore, the first issue of this chapter is to develop a task that facilitates the occurrence of action advantage. The second issue concerns the contributions of working memory components to following instructions in such a task environment. Finally, the third issue is to explore the extent to which working memory helps explain the rise of action advantage.

Action advantage is assumed to arise from a superior multimodal or imaginal-enactive representation for actions compared to a verbal-based representation for oral repetition (Allen, 2009; Gathercole, et al., 2008; Koriat, et al., 1990). This multimodal representation allows access to information from various dimensions, whereas the verbal representation is constrained by sequential processing. After a comparison with studies that have obtained action advantage (Allen, 2009; Gathercole, et al., 2008), it is inferred that the simplicity of the setting in the computer-based task might have overshadowed the benefit of the multimodal representation. Therefore, it is worth examining two tasks that have obtained action advantage, namely, the laminated-card task (Allen, 2009; Allen & Gathercole, 2008) and the classroom instructional task (Gathercole, et al., 2008).

The laminated card task contained a series of different actions on four geometric shapes, such as ‘push the cross, spin the star, drag the arch, touch the square’. The classroom task is a span task involving actions on a subset of six colourful objects and six colourful containers; for instance, an instruction with four actions is ‘pick up the yellow ruler and put it in the red box then pick up the blue rubber and put it in the yellow box’.

These two instructional tasks shared two commonalities that differed from the computer-based instructional task. First, both tasks were completed in a three-dimensional world involving the manipulation of objects by hand, which might provide proprioceptive feedback that would help speed up preparations for actions (Jeannerod, 1997b). The indirect contact with objects in the computer-based task might also demand extra work in the mapping of movements of an intermediate device (the mouse) on the computer screen, thus impairing the benefit gained from direct contacts with objects. Secondly, both the laminated card task and the classroom instructional task used natural sentences as spoken instructions, whereas instructions in the computer-based task were unnatural word-by-word sequences with regular intervals, which lacked the natural intonation and coherence of a natural sentence.

An important characteristic of the laminated card task is its variety of independent actions, which was different from the other two instructional tasks which contained fewer types of actions. Research has shown that action words automatically activate corresponding motor and premotor cortex (Hauk, Johnsrude, & Pulvermüller, 2004; Pulvermüller, Shtyrov, & Ilmoniemi, 2005), suggesting that action-based instructions by speech are likely to be encoded in an action form. Therefore, increased types of actions should be easily mapped onto a multimodal representation for future action recall, but might cause difficulty in describing these movements in words, thus facilitating the likelihood of action advantage.

The unique feature of the classroom instructional task was the involvement of colourful stationery objects that are common in daily life. Compared to the abstract geometric shapes used in the other two tasks, these stationery objects contained more information related to utility (Gibson, 1977, 1986). Specifically, people have plenty of experience using these common everyday objects; the presence of these objects therefore tended to activate the associated actions, and facilitate the preparation and execution of these actions.

As can be seen, both the laminated card task and the classroom instructional task contained rich cues beneficial for a multimodal representation. At the moment, it is difficult to select the most important factor; therefore all potential contributors designing a rich environment for the instructional task should be included. Specifically, this task would involve various types of actions upon colourful objects in a three-dimensional world, the so-called 3D instructional task. In this task, participants listened to the instructional sentences and recalled them either by oral repetition or through operations by hand. It is predicted that an action advantage should be obtained in a rich environment like this.

To investigate the involvement of working memory components in following spoken instructions in this rich task environment, the same dual task paradigm was used as in the last chapter. Previous experiments showed the significant involvement of the phonological loop and the central executive in following instructions in the computer-based instructional task. Therefore, the two working memory components were first examined in this 3D instructional task environment, with the articulatory suppression task and the backward counting task respectively. Therefore, three types of conditions were formed: the articulatory suppression condition, the backward counting condition, and a baseline condition serving as a comparison condition for articulatory suppression. As in Experiment 1, the articulatory suppression effect was the deduction in performance of recall in the articulatory suppression condition when compared to the baseline condition. The backward counting effect was the difference between the accuracy of recall in the backward counting condition and in the articulatory suppression condition, representing the specific disruption by the demand of the central executive after excluding the maintenance function of the phonological loop.

In summary, four hypotheses were tested in this experiment. First, in a rich task environment containing many facilitators for the rise of action advantage, the performance of recall by actions is expected to be superior to the oral repetition of instructions. Second, as the phonological loop supports the maintenance of spoken instructions, the concurrent articulatory suppression task during encoding should lead to impairment in recall. The third hypothesis relates to the roles of the central executive, which is assumed to be involved in many aspects of following instructions. For example, the central executive is hypothesized to be involved in directing eye movement during the searching for objects when their names are heard, thus helping maintain the sequence of to-be-enacted objects, which should be reflected in the increment of order errors as a result of backward counting. Another role of the central executive probably relates to binding elements in an action, as has been shown in the literature (Brown & Brockmole, 2010). This was tested by investigating whether there is a greater backward counting effect in the recall of combined elements than in individual elements. Given the responsibilities of the central executive assumed above, a substantial backward counting effect on memory of instructions is expected. Moreover, based on the previous findings and the versatility of the central executive, the contribution of the central executive is expected to be greater than that of the phonological loop. Therefore, the fourth hypothesis is that the backward counting effect is expected to be larger than the articulatory suppression effect. At present, whether the action advantage would be obtained remains uncertain, thus the interaction between working memory and recall is thus hard to predict. For this reason, no specific hypothesis regarding the contribution of working memory in different types of recall is made in this experiment.

Method

Participants

Twenty-four native English speakers at the University of York were recruited through the electronic booking system in exchange for course credit or an honorarium of £6. None had taken part in Experiments 1 and 2. There were 23 females and 1 male, aged from 18 to 26, with a mean age of 19.45.

Materials

The three-dimensional task environment involved colourful objects, including six small objects (a yellow ruler, a blue ruler, a white eraser, a green eraser, a red pencil and a black pencil), and six containers (a black box, a red box, a yellow bag, a white bag, a blue folder and a green folder). There were four types of actions, including 'touch', 'push', 'spin', and 'pick up…then put it into'. The action 'touch' was a gentle tap on the object, 'push' referred to shoving the object forward for a few centimetres, 'spin' was to make the object revolve on its own axis for a single turn, and 'pick up…then put it into…' were two concatenated actions requiring moving an object into a container. Similar to the 'drag… onto…' action in Experiment 1, the action 'pick up…then put it into…' was also considered as two actions because participants might pick up the correct object but put it into the wrong container.

Each instruction sentence involved five actions connected using the conjunction word 'and'. In an instructional sentence, there was no repetition of the same object, and adjacent objects were always in different colours. An example of a typical instruction sentence would be, 'Push the black pencil, and spin the green eraser, and touch the red pencil, and push the blue ruler, and touch the white eraser'.

All instructions were read by a native British female speaker with a clear pronunciation. Each instruction was read with the normal prosody of a common oral command and recorded as a whole sentence. The average duration of an instructional sentence was 9.22 s, ranging from 8.80 s to 9.66 s. This variation was due to the different numbers of words (24, 27 or 29) in the instructional sentences. Three lists of instructions were created. Each list contained 14 instructional sentences (2 practice trials and 12 formal trials) (see Appendix 3).

A total of 84 different three-digit numbers for the articulatory suppression condition and backward counting condition were randomly generated using Microsoft Excel software (see Appendix 4). These digits were also read and recorded by the same British female speaker, and the duration of a three-number digit was 3 seconds.

All objects were laid out on a 146cm (length) × 75cm (width) × 71cm (height) desk within an arm's reach of an adult (see Figure 3.1). Rather than changing locations between trials, the locations of objects remained the same throughout the experiment to decrease the effort of visual search in this rich environment.

Design

In a 3 × 2 mixed design, concurrent task was set as a within-subject variable including baseline, articulatory suppression and backward counting conditions. Recall type was a between-subject variable, including verbal and action recall condition. The main dependent variable was the mean number of correct actions per instruction sequence. Other measurements included elements (movement, colour, and object), combined elements (colourful object), and also the percentage of correct trials in each serial position.

Figure 3.1 The visual display of the 3D instructional task in Experiment 3.

Procedure

The experiment was carried out in a quiet room. Upon arrival, a participant was randomly assigned to one of the recall groups, and was then introduced to the instruction task. Before the formal test, participants finished a practice with six trials, two for each concurrent task condition; specifically, participants practiced the speed of counting in the articulatory suppression and backward counting task, at a rate of three digits every two seconds.

All participants sat in the middle front of a desk, facing the display of objects. The experimenter sat at the other desk 30 cm away from the participants, controlling the delivery of instructions. All spoken instructions were played through two speakers on the experimenter's desk which faced the participant. The volume was set at the appropriate level preferred by each participant. The verbal report of participants was recorded on the computer and performances in action recall conditions were captured by a camera. The experimenter also kept a written record of the recall.

All participants finished three conditions counterbalanced in order, namely, the baseline, articulatory suppression, and backward counting condition. The first two trials in each condition were practice ones. In the baseline conditions, the experimenter signalled the participant to prepare for the coming trial, and then triggered the playback of the instructions. The participant listened to the instruction, which was followed by a one second delay and a beep sound, indicating the start of recall. Based on the group he or she had been assigned to, the participant either repeated the instruction back to the experimenter (verbal recall) or performed the actions (action recall). Participants were told that they could either include or omit the conjunction words 'and' and 'then', which would not be counted in the scores.

The procedures in the articulatory suppression conditions were similar to the baseline conditions except that a participant first heard a three-digit number lasting three seconds, and began repeating the numbers continuously at a rate of three digits every two seconds. After a further three seconds of repeating the numbers, the instruction began to play. The participant continued repeating the three-digit number aloud while listening to the instructions until the beep sound. The procedure in the backward counting condition was similar to that in the articulatory suppression condition, except that the participants decreased the three-digit number by two continuously. The backward counting task was made easier comparing to decreasing the number by three in Experiment 1; this was done to prevent the floor effect that might be caused by the expected increased difficulty in the 3D instructional task. Any strategies employed by participants were investigated at the end of each condition, using a single question, 'if you are using any strategy, please state this'.

Results

Actions

The serial recall of actions was scored by averaging the correct actions per instruction across twelve formal trials. An action was considered correct only if the combination of movement, colour, shape, and its serial position were all correct. There were five actions in an instruction; therefore the scores of actions ranged from 0 to 5. The means and the standard deviations of actions are displayed in Table 3.1.

Table 3.1 Means (and standard deviations) of actions in Experiment 3

	Verbal recall	Action recall	Total
Baseline	2.95 (0.79)	3.86 (0.62)	3.41 (0.84)
Articulatory suppression	2.69 (0.78)	3.58 (0.61)	3.13 (0.82)
Backward counting	1.35 (0.76)	1.94 (0.71)	1.65 (0.78)
Means	2.33 (0.61)	3.12 (0.53)	2.73 (0.69)

A 3 × 2 (Concurrent task × Recall type) ANOVA showed significant main effect of concurrent task, F (2, 44) = 75.192, $p < 0.001$, $\eta_p^2 = 0.774$, MSE = 0.287, significant main effect of recall, action recall was better than verbal recall, F (1, 22) = 11.814, $p = 0.002$, $\eta_p^2 = 0.349$, MSE = 0.323. There was no significant interaction between concurrent task and recall type, F (2, 44) = 0.692, $p = 0.506$, $\eta_p^2 = 0.030$, MSE = 0.287.

Planned contrast indicated approaching significant articulatory suppression effect, F (1, 22) = 3.999, $p = 0.058$, $\eta_p^2 = 0.154$, MSE = 0.447, and significant backward counting effect, F (1, 22) = 99.633, $p < 0.001$, $\eta_p^2 = 0.819$, MSE = 0.533, and none of the two effects interact with type of recall (*ps* > 0.05).

The comparison of the two effects were examined with a 2 × 2 (Effect × Recall type) ANOVA. The larger backward counting effect was significantly larger than the articulatory suppression effect, F (1, 22) = 100.620, p < 0.001, $\eta_p{}^2$ = 0.821, MSE = 0.370. There was no significant main effect of recall type, F (1, 22) = 0.430, p = *0.519*, $\eta_p{}^2$ = 0.019, MSE = 0.305, and no significant interaction between effect and recall type, F (1, 22) = 0.836, p = 0.371, $\eta_p{}^2$ = 0.037, MSE = 0.370.

One-tailed dependent *t*-tests with Bonferroni corrections showed that the backward counting effect was significant in both verbal and recall groups (*ps* < 0.01). One-tailed independent *t*–tests with Bonferroni corrections found action advantage existed in baseline, articulatory suppression condition (*ps* < 0.05) but not in backward counting condition (p = 0.096).

Elements

In this experiment, each action contained three elements: movement, colour, and object. The scoring method was the same as that in Experiment 1; the recall of a colourful object was considered correct only when both colour and object were correctly paired together, and when it was also placed in the correct serial position. The chance levels were different for elements, with 20 percent for movement, 16.7 percent for colour, 16.7 percent for object, and 8 percent for colourful object; hence the scores of the elements were not compared. The means and standard deviations of elements and colourful objects are presented in Table 3.2.

A 3 × 2 × 3 (Concurrent task × Recall type × Element) ANOVA was conducted. There was significant main effect of concurrent task, F (2, 44) = 94.516, p < 0.001, $\eta_p{}^2$ = 0.811, MSE = 0.649, recall type, F (1, 22) = 11.718, p = 0.002, $\eta_p{}^2$ = 0.348, MSE = 0.254, and element, F (1.31, 28.81) = 62.195, p < 0.001, $\eta_p{}^2$ = 0.739, MSE = 0.124. There was significant interaction between concurrent task and recall type, F (2, 44) = 0.073, p < 0.929, $\eta_p{}^2$ = 0.003, MSE = 0.649, and between concurrent task and element, F (2.77, 60.83) = 16.636, p < 0.001, $\eta_p^{\ 2}$ = 0.431, MSE = 0.052, but no other two-way or three-way interaction (*ps* > 0.05).

Table 3.2 Means (and standard deviations) of elements in Experiment 3

		Verbal recall	Action recall	Total
Movement	Baseline	3.28 (0.74)	4.04 (0.54)	3.66 (0.74)
	Articulatory suppression	3.16 (0.71)	3.84 (0.52)	3.49 (0.70)
	Backward counting	2.09 (0.76)	2.51 (0.71)	2.30 (0.75)
	Means	2.84 (0.61)	3.46 (0.45)	3.15 (0.62)
Colour	Baseline	3.84 (0.57)	4.59 (0.37)	4.22 (0.61)
	Articulatory suppression	3.70 (0.84)	4.40 (0.38)	4.05 (0.73)
	Backward counting	1.95 (0.86)	2.74 (0.70)	2.35 (0.87)
	Means	3.16 (0.63)	3.91 (0.36)	3.54 (0.63)
Object	Baseline	3.94 (0.57)	4.69 (0.31)	4.31 (0.59)
	Articulatory suppression	3.78 (0.87)	4.50 (0.30)	4.14 (0.74)
	Backward counting	2.15 (0.85)	2.89 (0.73)	2.52 (0.86)
	Means	3.29 (0.65)	4.03 (0.33)	3.66 (0.63)
Colourful object	Baseline	3.78 (0.61)	4.58 (0.38)	4.18 (0.64)
	Articulatory suppression	3.58 (0.91)	4.38 (0.37)	3.98 (0.80)
	Backward counting	1.82 (0.86)	2.63 (0.75)	2.22 (0.89)
	Means	3.06 (0.65)	3.87 (0.37)	3.46 (0.66)

Planned contrasts indicated no significant articulatory suppression effect, F (1, 22) = 1.990, p = 0.172, $\eta_p^{\ 2}$ = 0.083, MSE = 0.349, but significant backward counting effect, F (1, 22) = 145.082, p < 0.001, $\eta_p^{\ 2}$ = 0.868, MSE = 0.375, which was more disruptive to the memory for colour and object than to the memory of movement (*ps* < 0.001).

One-tailed dependent *t*-tests with Bonferroni corrections showed no significant articulatory suppression effect in any of the elements and colourful object (*ps* > 0.05), and significant backward counting effect in all elements and colourful object (*ps* <

0.01). One-tailed independent *t*-tests with Bonferroni corrections showed that action advantage existed in all elements and also colourful objects ($ps < 0.05$).

Binding

Although the result in Experiment 1 indicates no contributions from the central executive to the binding of movement to object, the result may be different in a task environment containing rich cues. Nevertheless, the method of testing binding was similar as that in Experiment 1.

There were three types of elements, movement, colour and object. Two types of binding were tested. One is to bind colours to objects, the successful binding led to correct colourful objects; therefore, the cost of this binding was reflected in lower scores for objects compared to colorful objects. The other type of binding is the binding of movements and colourful objects to form correct actions, and its cost was indicated as the lower scores for colourful objects compared to actions.

If central executive was truly involved in binding, then the backward counting should be more disruptive to bound entities than to elements. That is, backward counting effect should be significantly larger in colorful object than in object, and it should also be larger in action than in colourful object. To test whether the binding role of central executive differed in verbal and action recall, the variable recall was also included.

First, consider binding colour to the object. In a 2 × 2 × 2 (Backward counting × Binding × Recall type) ANOVA, the variable backward counting contained articulatory suppression and backward counting conditions, the variable binding included colourful object and object, and the variable recall included verbal and action recall. There was significant main effect of backward counting, $F\ (1, 22) = 157.660$, $p < 0.001$, $\eta_p{}^2 = 0.878$, MSE = 0.433, binding, $F\ (1, 22) = 58.186$, $p < 0.001$, $\eta_p{}^2 =$

0.726, MSE = 0.021, and recall type, F (1, 22) = 8.328, p = 0.009, $\eta_p{}^2$ = 0.275, MSE = 0.429. The interaction between backward counting and binding was significant, with a greater backward counting disruption to the memory for colourful object than to the memory for object, F (1, 22) = 6.392, p = 0.019, $\eta_p{}^2$ = 0.225, MSE = 0.015. There was no other two-way or three-way interaction (*ps* > 0.05).

Second, consider binding movement and colourful object. This was tested by a 2 × 2 × 2 (Backward counting × Binding × Recall type) ANOVA. The variable binding contained colourful object and action. Results showed significant main effect of backward counting, F (1, 22) = 131.200, p < 0.001, $\eta_p{}^2$ = 0.856, MSE = 0.479, binding, F (1, 22) = 174.799, p < 0.001, $\eta_p{}^2$ = 0.888, MSE = 0.069, and recall type, F (1, 22) = 17.665, p < 0.001, $\eta_p{}^2$ = 0.445, MSE = 0.023. There was a significant interaction between backward counting and binding, but the backward counting disrupted colourful object more than action, F (1, 22) = 17.665, p < 0.001, $\eta_p{}^2$ = 0.445, MSE = 0.023. There was no other two-way interaction (*ps* > 0.05), but a significant three-way interaction, F (1, 22) = 6.427, p = 0.019, $\eta_p{}^2$ = 0.226, MSE = 0.023, that is, the greater backward counting effect in colourful object compared to that in action was significant in the verbal recall group, F (1, 11) = 24.941, p < 0.001, $\eta_p{}^2$ = 0.694, MSE = 0.021, but was not significant in the action recall group, F (1, 11) = 1.276, p = 0.283, $\eta_p{}^2$ = 0.104, MSE = 0.026.

Serial positions

Each position was coded for the percentage of correct trials, ranged from 0 to 1. A 3 × 2 × 5 (Concurrent task × Recall type × Serial position) ANOVA showed significant main effect of serial position, F (2.74, 60.33) = 30.162, p < 0.001, $\eta_p{}^2$ = 0.578, MSE = 4.245, concurrent task, F (2, 44) = 75.916, p < 0.001, $\eta_p{}^2$ = 0.775, MSE = 8.039, and recall type, F (1, 22) = 12.027, p = 0.002, $\eta_p{}^2$ = 0.353, MSE = 1.865. There was no two-way or three-way interaction (ps > 0.05).

The post-hoc tests with Bonferroni corrections showed significant decrement of performance of recall between positions 1 and 2 (p = 0.011), and positions 2 and 3 (p < 0.001), but there was no significant decrement between the other adjacent positions (ps > 0.05). One-tailed independent t-tests with Bonferroni corrections showed that the action advantage was present in position 1, 2, 3, and 5 (ps < 0.05) but not in position 4 (p = 0.118).

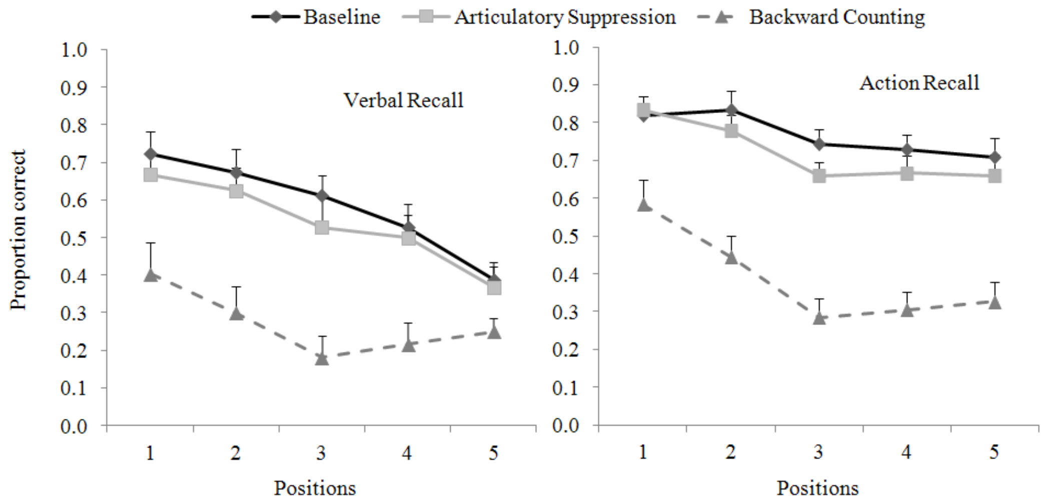

Figure 3.2 The serial position curves (means and standard errors) as functions of concurrent tasks and recall type in Experiment 3

Proportions of order errors

Order errors are the differences between the number of serial-recall actions and the correct actions regardless of order. Proportions of order errors per action recalled were computed by dividing the total number of order errors by the number of actions recalled regardless of order. The means and standard deviations as functions of concurrent tasks and recall type are shown in Table 3.3.

Table 3.3 Means (and standard deviations) of the proportion of order errors in Experiment 3

	Baseline	Articulatory suppression	Backward counting
Verbal recall	0.10 (0.11)	0.10 (0.11)	0.22 (0.16)
Action recall	0.03 (0.05)	0.04 (0.04)	0.22 (0.22)

A 3 × 2 (Concurrent task × Recall type) ANOVA was conducted. There was significant main effect of concurrent task, F (1.35, 29.69) = 10.816, p = 0.001, $\eta_p{}^2$ = 0.330, MSE = 0.026, but no significant main effect of recall type, F (1, 22) = 2.292, p = 0.144, $\eta_p{}^2$ = 0.094, MSE = 0.005. There was no significant interaction between concurrent task and recall type, F (1.35, 29.69) = 0.402, p = 0.592, $\eta_p{}^2$ = 0.018, MSE = 0.026. Post-hoc tests showed with Bonferroni corrections significantly less proportions of order errors in the articulatory suppression conditions than in the backward counting conditions (p = 0.009), and no significant difference between the articulatory suppression and the baseline conditions (p > 0.05).

Strategy report

Among the twenty-four participants, fifteen participants reported using strategies, eight in the verbal recall group and seven in the action recall group. The strategies can be summarized as six categories (see Table 3.4). For each given strategy, the numbers of responders were counted for each recall group; this is the count score. The percentage score for a given strategy represents the percentage of participants who indicated using that strategy. It was calculated by dividing the count score with the total number of responders in that recall group. The most commonly-used strategy was the visual tracking of the objects as their names were mentioned during encoding, indicated by most participants in both verbal and action recall. Both groups of participants also indicated imagining themselves performing the instructed actions during encoding.

Table 3.4 Self-report strategies in Experiment 3

	Verbal recall (N=8)		Action recall (N=7)		Total (N=15)	
	Count	Percent	Count	Percent	Count	Percent
Visual tracking	7	88%	7	100%	14	93%
Imagining carrying out the actions	3	38%	3	43%	6	40%
Decreasing interference	1	13%	5	71%	6	40%
Grouping actions	2	25%	3	43%	5	33%
Binding elements	2	25%	0	0%	2	13%
Verbal rehearsal	0	0%	1	14%	1	7%

Note. There were participants indicating use of multiple strategies, therefore the total scores of percent cores in a column can exceed 100%

Discussion

Summary of main results

This experiment set out to develop a rich task environment that gives rise to action advantage and explores the role of working memory, specifically, the phonological loop and the central executive, in this new task paradigm of following instructions. Four main findings emerged from this experiment. First, in a task environment with rich cues, action advantage was obtained, in line with previous research (Allen, 2009; Allen & Gathercole, 2008; Gathercole, et al., 2008). Articulatory suppression caused marginally significant disruption on recall, and backward counting produced a large disruption to the performance of recall. Direct contrast of the two effects showed that the backward counting effect was significantly larger than the articulatory suppression effect. These results thus verified the four hypotheses proposed in the introduction. Although there was evident action advantage, there was no interaction between working memory and recall, suggesting similar contributions of working memory to the two types of recall.

Action advantage

Comparing to the computer-based instructional task, an environment with rich cues finally gave rise to the action advantage. The richness of the task environment reflected in the increased types of dimensions, such as adding the dimension of colour, as well as in the variations inside a dimension like the increased types of actions. The increased variations also led to the accrued possibilities of combinations that contribute to the richness of environment. Moreover, the everyday objects contained rich semantic information relating to the affordances of these objects, which were likely to prime appropriate actions (Craighero, Fadiga, Rizzolatti, & Umiltà, 1998; Craighero, Fadiga, Umiltà, & Rizzolatti, 1996). Furthermore, the depth perception in this three-dimensional task was lacked in the computer-based task. All these factors, and more likely, their aggregated effects might have contributed to this action advantage.

Indeed, action advantage not only existed in recall of actions, but also in memory of all dimensions, movement, colour and object – and in the combinations of colour and object, as well as in all serial positions. This implies that the preparation for action recall enhanced all aspects of elements and sequential information, and together, these increased the performance of action recall and gave rise to the action advantage.

Specifically, according to the dual representation account (Koriat, et al., 1990), a superior action-based representation was formed for actions in contrast to a verbal representation for repetition. In the 3D instructional task, the richness of cues might have benefited the formation of a multimodal representation for actions, but caused cognitive demand on a sequentially-organized verbal representation for oral repetition, leading to the action advantage. If this assumption is true, working memory should show different contributions to the two representations, which would be reflected as a significant interaction between concurrent tasks and recall. The phonological loop should be more involved in verbal representation for oral repetition, whereas the central executive should be more needed in binding elements in a multimodal representation for actions. However, the results of this experiment showed no such interaction between concurrent tasks and recall; both articulatory suppression and backward counting disrupted the two types of recall similarly, suggesting similar contributions from the phonological loop and the central executive to the two types of recall. In other words, neither the phonological loop nor the central executive is likely to be the source of action advantage.

One possibility for the rise of action advantage may be related to the output interference, which can be inferred based on the contrasting shapes of serial position curves of the verbal and action recall conditions (see Figure 3.2). Take baseline condition for example, the curve has a steep slope in verbal recall condition whereas in the action recall condition, the line is relatively flat. Specifically, compared to the action recall condition, a larger primacy effect can be observed in the verbal recall condition. This implies growing proactive verbal output interference, that is, the item being recalled interferes with the items not yet recalled (Cowan, et al., 2002).

The contributions of the working memory

First consider the phonological loop. The articulatory suppression effect was approaching significance for the recall of actions but not individual elements. These results indicate that the support from the phonological loop in encoding and maintenance of spoken instructions was relatively weak. One possible reason for this is that the instructional sentences were longer in this experiment (24-29 words) than those in the computer-based task (18 words), thus exceeding the capacity of the phonological loop. The participants may therefore depend on other strategies, such as retaining most of the information in the visuospatial sketchpad. This interpretation is supported by subjective reports from over 90 percent of participants that they actively tracked the locations of objects as a means of memorizing the instructions.

This active tracking of objects in space is likely to be supervised under the attentional control of the central executive (Postle, et al., 2006). In this experiment, this argument is supported by the finding of a greater backward counting effect in memory of colour and object than in memory of movement, suggesting the active involvement of the central executive in memorizing the locations of objects, and from which, the detailed visual information was extracted during recall. The use of cues in the external world as extensions of memory is not uncommon. Compared with remembering all the detailed features of an object, it has been observed that people tend to encode the spatial locations of objects, and follow these deictic pointers during retrieval (Spivey, Richardson, & Fitneva, 2004). These deictic pointers can be especially useful when an object contains multiple features such that the memory load exceeds the capacity of working memory. In addition, when the time allowed for encoding each object is limited, these pointers can also be efficient and thus helpful.

The attentional role of the central executive in directing eye movements probably also served in encoding and maintenance of ordinal information of actions. In this experiment, backward counting effect was evident in the proportion of order errors. The increased errors in orders suggest that occupying executive resources caused great difficulty in maintaining the order of sequential operations upon different objects. Although maintaining sequence is usually the responsibility of the phonological loop, other strategies might be adopted when the length of instructions exceeds its capacity (Salamé & Baddeley, 1986). This explanation is consistent with finding that articulatory suppression has no effect on order errors in this experiment.

It is possible that using locations as memory cues may also facilitate the binding of the events in a specific location, such as binding visual features in a colourful object. In this experiment, the role of the central executive in binding was examined. The backward counting effect was larger in colourful objects than in objects alone; suggesting that binding colour to a corresponding object requires the central executive. However, there was no such trend in binding movement to a colourful object. In fact, the backward counting disrupted memory of colourful objects more than memory of actions, suggesting that binding movement to a colourful object places less of a demand on the central executive than retaining a bound entity of a colourful object. The two pieces of evidence in binding thus suggest that the central executive was involved in binding and maintaining the visual features of a colourful object, but did not facilitate the linking of movement to the related object. In the computer-based instructional task, the central executive did not help bind movement to a corresponding object. Nevertheless, forty percent of participants reported imagining themselves performing the actions during encoding; it is possible that it is the process of imagining oneself carrying out the actions that helped build the strong links between the actions and associated objects.

Next step

In this 3D task environment with rich cues, a superior recall through actions than through oral repetition was secured; however, the reason for the rise of action advantage remains speculative. Both the phonological loop and the central executive showed a similar contribution to the verbal and action recall, suggesting they were unlikely to be the sources of action advantage. Even from the attentional view of working memory (Cowan, 1999), disruption from concurrent tasks was considered to be an indication of attentional demand; the similar disruptive effect of the same concurrent task to the two types of recall also implies that an equal amount of attention is involved in representing instructions for oral repetition and actions. It seems unlikely, therefore, that the attentional aspect helps account for the action advantage.

The report of frequent eye tracking of objects during encoding suggests the possibility for the formation of a spatial representation for the actions to-be-performed. Together with the evidence of the weak support of the phonological loop, it is highly likely that the other working memory storage component, the visuospatial sketchpad, is involved in supporting the memory process of instructions. If this assumption is true, then preventing active eye tracking of objects in space or disrupting the accessibility of visual and spatial information should impair recall.

These two hypotheses were tested in Experiments 4 and 5. In Experiment 4, a spatial tapping task was used to disrupt the visuospatial sketchpad, and in Experiment 5, participants received instructions either with or without a view of the spatial array of objects.

Chapter 4

Exploring the visuospatial sketchpad in a rich environment

Introduction

The main purpose of this chapter is to investigate the role of the visuospatial sketchpad in supporting the following of instructions. In a rich task environment, the benefit of recalling instructions by actions than by oral repetition has been established. However, the underlying mechanism behind this action advantage remains unknown. The previous experiment showed that both the central executive and the phonological loop were involved in encoding spoken instructions but were unlikely to be the source of action advantage. It is suspected that the visuospatial sketchpad might have contributed greatly to this encoding process, especially to the forming of a map of the series of locations of to-be-enacted objects. It is thus hypothesized that the visuospatial sketchpad is involved in following spoken instructions.

Moreover, the occurrence of action advantage in a rich task environment but not in the computer-based task implies that the richness of these visual, spatial and motor cues contributed to the rise of action advantage. According to the multicomponent working memory model, the visuospatial sketchpad is in charge of encoding visual, spatial and motor-related information (Baddeley, 2000, 2007). The visuospatial sketchpad may contribute more to the formation of a multimodal representation for action recall compared to a verbal representation for repetition, and may therefore be the source of action advantage. If this is the case, disrupting the visuospatial sketchpad should impair the performance of actions more than oral repetition. In other words, an interaction between the visuospatial sketchpad and recall is expected, reflected as the deduction or the elimination of action advantage when the function of the visuospatial sketchpad is interfered with. Therefore, another purpose of this chapter is to explore the contribution of the visuospatial sketchpad to the action advantage.

In Experiment 4, a spatial tapping task was used to interfere with the spatial encoding supported by the visuospatial sketchpad. In Experiment 5, visuospatial information was blocked in such a way that participants received instructions either with or without a view of the spatial array of objects. As in the previous experiments, all concurrent tasks occurred only during the presentation of the instructions.

Experiment 4

Introduction

In the multicomponent working memory model (Baddeley, 2000), the visuospatial sketchpad is responsible for holding visual, spatial and motor information, and the relationship between the three subcomponents has been considered both separable (Klauer & Zhao, 2004) and interactive (Wager & Smith, 2003b) (for more information see discussion of the visuospatial sketchpad in Chapter 1). In particular, the distinction between the visuospatial sketchpad and the central executive is less clear; for example, compared with the simultaneous presentation of visuospatial stimuli, sequential visuospatial tasks tend to invoke the use of central executive resources (Fisk & Sharp, 2003; Jones, Farrand, Stuart, & Morris, 1995a; Klauer & Stegmaier, 1997; Rudkin, et al., 2007). Therefore, it is important to select a pure concurrent task that only disrupts the visuospatial sketchpad without placing demands on the central executive.

Among the many interfering tasks, the tapping task has been a frequently used task selectively interfering with the visuospatial sketchpad. The earliest version of the tapping task required continuous tapping of four separated metal plates positioned in a square arrangement, and participants had to tap in a clockwise direction as quickly as possible using a stylus, with feedback provided by four illuminated indicator lamps. This continuous tapping task showed selective disruption on spatial reasoning but not on verbal reasoning (Farmer, et al., 1986), suggesting that the task contained a distinctive visuospatial component. In a later study, the visual feedback was removed and this continuous tapping task showed specific interference in a spatial memory task but not on a movement memory task (Smyth, et al., 1988), suggesting that the task relies mainly on the spatial storage component. A revised version of the tapping task required participants to tap at a paced rate, around one tap per second (Salway & Logie, 1995). More complex versions of the tapping task used complicated tapping patterns, such as tapping in specified patterns within a 3×3 or 5×5 matrix (Coluccia, Bosco, & Brandimonte, 2007; Garden, Cornoldi, & Logie, 2002; Klauer & Zhao, 2004), or tapping in a figure of eight (Allen, et al., 2009; Deyzac, Logie, & Denis, 2006; Pearson, Logie, & Gilhooly, 1999; Postle, et al., 2006). These complex tapping tasks were also found to load the spatial component in the visuospatial sketchpad. Other tapping tasks, such as paced or syncopated tapping tasks, required the tapping of a single key without any spatial demand. The syncopated tapping task was found to disrupt verbal serial recall (Hall & Gathercole, 2011; Larsen & Baddeley, 2003), possibly due to the sharing of the same speech motor program of rehearsal in the phonological loop (Saito, 1994).

Among these variations, a relatively pure spatial tapping task involving tapping four spatial locations at a paced rate was chosen, which is supposed to cause selective interference to the spatial working memory. Given that most participants in previous experiments have reported tracking the locations of objects in this rich 3D task environment, interfering with this spatial coding should impair the memory of locations, and consequently impair the memory of instructions. Moreover, as mentioned before, interfering with spatial coding is likely to disrupt the multimodal representation for action recall, but have little effect on the verbal-based representation for verbal repetition, leading to the decrement of action advantage. Therefore, a spatial tapping task should impair the performance of recall of instructions, and also decrease or eliminate the action advantage.

A further comparison made in this study was between the contributions of the phonological loop and the visuospatial sketchpad to following spoken instructions. The articulatory suppression task was changed to repeating '1-2-3-4' at the same rate as that in tapping, to equate as closely as possible the demands of the two activities. The three concurrent task conditions were therefore tapping, articulatory suppression condition, and no activity.

In summary, two hypotheses were tested in this experiment. First, that encoding visual, spatial and motor information is important in following instructions in a rich task environment, which is likely to be supported by the visuospatial sketchpad; therefore, disrupting its functioning by a spatial tapping task would impair the performance of recall. Second, spatial coding is assumed to contribute to the action advantage; hence the tapping effect was expected to be greater in the action recall than that in the verbal recall.

Method

Participants

Twenty-four native English speakers at the University of York were recruited through the electronic booking system in exchange for course credit or an honorarium of £6. None had taken part in the previous three experiments. There were 19 females and 5 males, aged from 18 to 27, with a mean age of 19.71.

Materials

The instructional materials were similar to those in Experiment 3, except that all four types of actions were included in each sentence without repetition of movements. Each instructional sentence contained five actions, a typical example being, 'pick up the yellow ruler then put it into the red box and spin the blue ruler and touch the white rubber and push the black box'.

Three lists of instructions were constructed. Each list included fourteen instructional sentences, of which the first two were practice trials and the other twelve were formal trials (see Appendix 5). The numeric keypad of a standard Dell keyboard was adapted into a four-key device, with four keys at the corners (numbered 7, 9, 1, 3) and all the other keys removed. The keypad was hidden from the view of participants by attaching it to the top of a computer case placed under the desk near the dominant-hand side of the participant.

The visual display of objects was similar to that in Experiment 3 except for two changes. One change was that the objects in this experiment were arranged so that they were spread out across a larger space. This change not only avoided the cluster of objects in a small space, but also increased the variability of the spatial information, thus facilitating the use of the visuospatial sketchpad. The other change was that the visual display was changed between trials in order to ensure that each trial resembled a new learning environment. However, within each individual trial, that is, during the encoding and recall stage, the visual display remained the same.

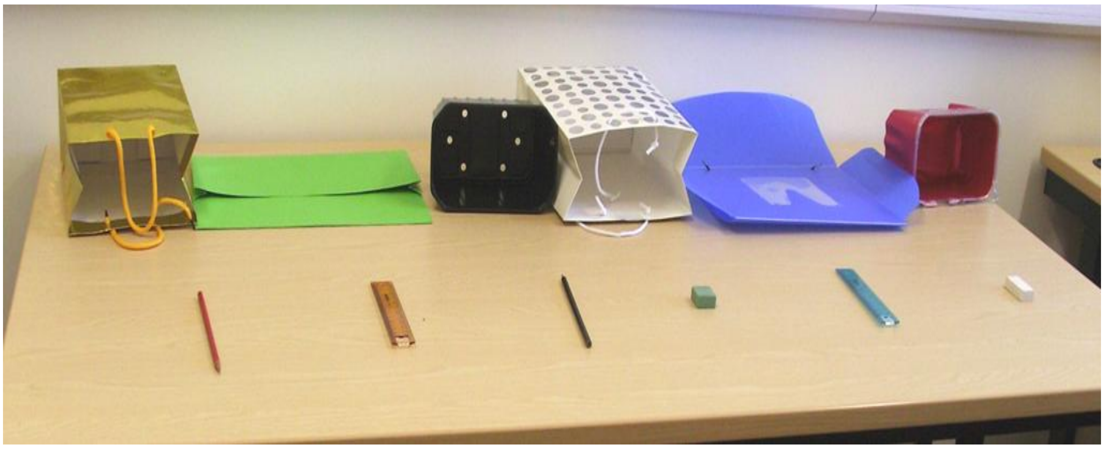

Figure 4.1 Visual display of 3D instructional task in Experiment 4

Design

In a 3×2 mixed design, concurrent task was set as the within-subject variable including baseline, articulatory suppression and tapping condition; and recall type was a between-subject variable, including verbal and action recall. The main dependent variable was the mean number of correct actions per instruction sequence.
Other measurements included elements (movement, colour, and object), combined elements (colourful object), and also the percentage of correct trials in each serial position.

Procedure

Participants signed the consent forms and were randomly assigned to the verbal or action recall group. They were then introduced to the instruction task, and were given a practice task containing six trials, two for each concurrent task condition. The pace of articulation and tapping were also practiced.

Each participant finished three conditions: the baseline, articulatory suppression, and tapping conditions. The sequence of the three conditions was counterbalanced using a Latin-square design, and the lists were rotated to equate the probability of different combinations between conditions and lists. The baseline conditions were same as those in Experiment 3.

In the articulatory suppression conditions, the participant first heard the sound of the words '1-2-3-4' which lasted three seconds, after which he or she repeated this aloud at the same rate continuously. After a further three seconds, the participant heard the spoken instructions while repeating '1-2-3-4' throughout the delivery of the instructions and the one-second gap until the beep sound, and then began to recall.

In the tapping condition, upon hearing the command sound 'start', the participant began to tap the four keys 1-7-9-3 clockwise on the keypad at a pace of three seconds per circle. After a further three seconds, the participant heard the instructional sentences while continuing tapping until the beep sound. Participants were told to use only their forefingers to tap.

After the beep sound, the participant either repeated the instruction back (verbal recall) or performed the actions (action recall) according to the group he or she had been assigned to. Participants were allowed to omit the conjunction words 'and' and 'then', but they could also choose to include them. These conjunctional words were not counted in the scores of verbal recall. At the end of each trial, the experimenter randomly changed the locations of three objects on the table; participants were asked to close their eyes during the change.

The recording method was same as that in Experiment 3. Any strategies used by participants were investigated at the end of each condition using a single question, 'if you are using any strategy, please state this'.

Results

Actions

The serial recall of actions was scored by averaging the correct actions per instruction across twelve formal trials. The scoring method was same as that in Experiment 3, and the means and the standard deviations of actions as functions of concurrent tasks and recall type were displayed in Table 4.1.

Table 4.1 Means (and standard deviations) of actions in Experiment 4

	Verbal recall	Action recall	Means
Baseline	3.44 (0.79)	4.11 (0.87)	3.78 (0.88)
Articulatory suppression	3.10 (0.81)	3.65 (0.80)	3.37 (0.83)
Tapping	2.98 (1.18)	3.89 (1.01)	3.44 (1.17)
Means	3.18 (0.80)	3.88 (0.80)	3.53 (0.86)

A 3 × 2 (Concurrent task × Recall type) ANOVA revealed significant main effect of concurrent task, F (2, 44) = 3.690, p = 0.033, $\eta_p{}^2$ = 0.144, MSE = 0.305, significant main effect of recall type, with a better performance of recall by actions than by oral repetition, F (1, 22) = 4.695, p = 0.041, $\eta_p{}^2$ = 0.176, MSE = 0.641. The interaction between concurrent task and recall type was not significant, F (1, 22) = 0.618, p = 0.544, $\eta_p{}^2$ = 0.027, MSE = 0.305.

Planned contrasts found significant articulatory suppression effect, F (1, 22) = 7.214, p = 0.013, $\eta_p{}^2$ = 0.247, MSE = 0.543, but no significant tapping effect, F (1, 22) = 3.537, p = 0.073, $\eta_p{}^2$ = 0.139, MSE = 0.773. And the two effects did not interact with the recall type (ps < 0.05).

Individual scores of the articulatory suppression and tapping effect were calculated. A 2 × 2 (Effect × Recall type) ANOVA showed no significant difference between the two effects, F (1, 22) = 0.207, p = 0.654, $\eta_p{}^2$ = 0.009, MSE = 0.258, no significant main effect of recall type, F (1, 22) = 0.028, p = 0.868, $\eta_p{}^2$ = 0.001, MSE =

0.529, and no significant interaction between effect and recall type, F (1, 22) = 1.424, p = 0.245, η_p^2 = 0.061, MSE = 0.258.

One-tailed dependent t-tests with Bonferroni correction showed no significant articulatory suppression effect in the verbal and action recall group (ps > 0.05). One-tailed independent t-tests with Bonferroni corrections failed to show significant action advantage in the baseline, articulatory suppression, and tapping conditions (ps > 0.05).

Elements

Each action contained three elements: movement, colour and object. The scoring method for these elements was same as that in Experiment 3. As in the previous experiments, the chance levels of the different elements varied (see Experiment 3 for details). The means and standard deviations of the elements and colourful objects as functions of concurrent tasks and the type of recall are presented in Table 4.2.

A 3 × 2 × 3 ANOVA (Concurrent task × Recall type × Element) was conducted. The result showed significant main effect of concurrent task, F (2, 44) = 3.633, p = 0.035, η_p^2 = 0.142, MSE = 0.586, element, F (1.26, 27.72) = 49.329, p < 0.001, η_p^2 = 0.692, MSE = 0.116, but no significant main effect of recall type, F (1, 22) = 3.269, p = 0.084, η_p^2 = 0.129, MSE = 0.461. There was a significant interaction between concurrent task and element, F (2.81, 61.81) = 3.022, p = 0.039, η_p^2 = 0.121, MSE = 0.055, but there was no other two-way or three-way interaction (ps > 0.05).

Planned contrasts showed significant articulatory suppression effect, F (1, 22) = 6.285, p = 0.020, η_p^2 = 0.222, MSE = 0.303, and significant tapping effect, F (1, 22) = 4.305, p = 0.050, η_p^2 = 0.164, MSE = 0.541. Neither effect interacted with elements (ps > 0.05). Nevertheless, direct contrast of the two effects shows an interaction, which was mainly reflected as a relatively larger articulatory suppression effect compared with the tapping effect on memory of movement (p < 0.001), whereas tapping was found to be more disruptive than articulatory suppression on memory of colour and object ($p < 0.001$).

One-tailed dependent t-tests with Bonferroni corrections found significant articulatory suppression effect in movement ($p = 0.028$), but not in colour, object nor in colourful object ($ps > 0.05$). Tapping effect was not significant in any of the elements ($ps > 0.05$). One-tailed independent t-tests with Bonferroni corrections showed that the action advantage was absent in all types of elements ($ps > 0.05$).

Table 4.2 Means (and standard deviations) of elements in Experiment 4

		Verbal recall	Action recall	Means
Movement	Baseline	3.71 (0.66)	4.20 (0.83)	3.95 (0.78)
	Articulatory suppression	3.32 (0.75)	3.85 (0.75)	3.59 (0.77)
	Tapping	3.34 (1.12)	4.09 (0.95)	3.72 (1.09)
	Means	3.46 (0.73)	4.05 (0.76)	3.75 (0.79)
Colour	Baseline	4.07 (0.74)	4.54 (0.58)	4.31 (0.70)
	Articulatory suppression	3.93 (0.46)	4.28 (0.76)	4.06 (0.65)
	Tapping	3.72 (0.94)	4.19 (0.90)	3.95 (0.93)
	Means	3.91 (0.62)	4.34 (0.69)	4.11 (0.67)
Object	Baseline	4.10 (0.70)	4.62 (0.59)	4.36 (0.69)
	Articulatory suppression	3.88 (0.58)	4.36 (0.71)	4.12 (0.68)
	Tapping	3.75 (1.02)	4.28 (0.88)	4.01 (0.97)
	Means	3.91 (0.66)	4.42 (0.68)	4.16 (0.71)
Colourful object	Baseline	4.00 (0.76)	4.53 (0.60)	4.26 (0.72)
	Articulatory suppression	3.83 (0.55)	4.27 (0.76)	4.05 (0.68)
	Tapping	3.63 (1.02)	4.18 (0.90)	3.90 (0.98)
	Means	3.82 (0.68)	4.33 (0.69)	4.07 (0.72)

Binding

The method of testing binding was the same as that in Experiment 3 except that in this experiment, it was the role of visuospatial sketchpad in binding elements being examined. The tapping effect on binding a colour to an object was tested using a 2 × 2 × 2 (Tapping × Binding × Recall type) ANOVA, the variable tapping included baseline and tapping condition, and the variable binding included colourful object and object. Results showed significant main effect of tapping, F (1, 22) = 4.826, p = 0.039, η_p^2 = 0.180, MSE = 0.003, binding, F (1, 22) = 19.173, p < 0.001, η_p^2 = 0.466, MSE = 0.013, but no significant main effect of recall type, F (1, 22) = 3.291, p = 0.083, η_p^2 = 0.803, MSE = 0.519. There was no significant interaction between tapping and binding, F (1, 22) = 0.073, p = 0.789, η_p^2 = 0.003, MSE = 0.013, and no other two-way or three-way interaction (*ps* > 0.05).

The tapping effect on binding a movement to corresponding colourful object was tested using a 2 × 2 × 2 (Tapping × Binding × Recall type) ANOVA, and the variable binding included colourful object and action. Results showed significant main effect of tapping, F (1, 22) = 4.159, p = 0.054, η_p^2 = 0.159, MSE = 0.699, binding, F (1, 22) = 60.660, p < 0.001, η_p^2 = 0.7341, MSE = 0.090, and recall type, F (1, 22) = 4.242, p = 0.051, η_p^2 = 0.162, MSE = 0.621. There was no significant interaction between binding and tapping, F (1, 22) = 0.162, p = 0.691, η_p^2 = 0.007, MSE = 0.016, and no other two-way or three-way interaction (*ps* > 0.05).

Serial positions

Each position was coded for the percentage of correct trials, ranged from 0 to 1. A 3 × 2 × 5 (Concurrent task × Recall type × Serial position) ANOVA found significant main effect of serial position, F (2.39, 52.62) = 38.669, p < 0.001, η_p^2 = 0.637, MSE = 4.483, concurrent task, F (2, 44) = 4.175, p = 0.022, η_p^2 = 0.160, MSE = 8.933, and approaching significant main effect of recall type, F (1, 22) = 4.092, p = 0.055, η_p^2 = 0.157, MSE = 3.911. There was a significant interaction between serial position and recall type, F (2.39, 52.62) = 9.245, $p < 0.001$, $\eta_p^2 = 0.296$, MSE = 4.483, and no other two-way or three-way interaction (*ps* > 0.05).

The post-hoc tests with Bonferroni corrections found the decrement of recall was significant between positions 1 and 2 ($p = 0.014$), positions 2 and 3 ($p = 0.003$), positions 3 and 4 ($p = 0.001$), but not between positions 4 and 5 ($p = 1.000$). These effects are presented in Figure 4.2.

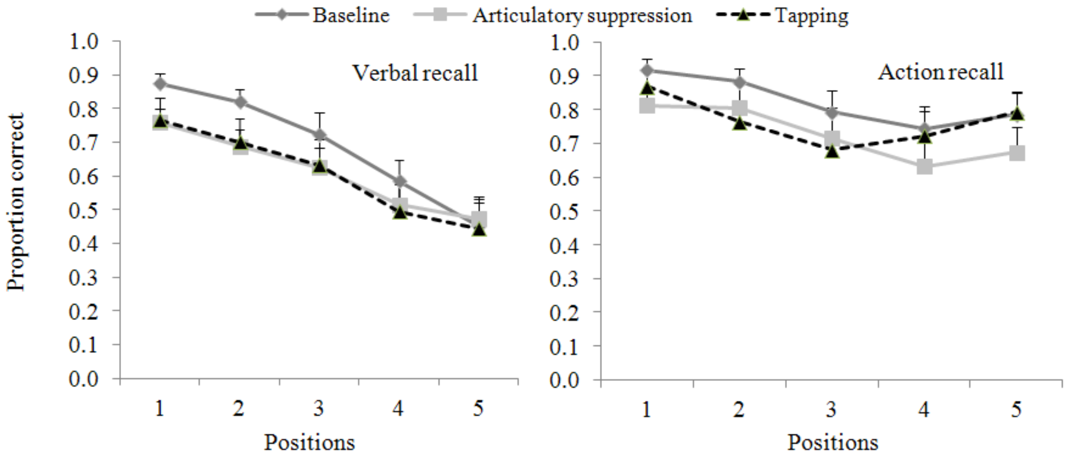

Figure 4.2 The serial position curves (means and standard errors) as functions of concurrent tasks and types of recall in Experiment 4.

Proportions of order errors

The scoring method of the proportions of order errors was same as that in Experiment 3. The means and standard deviations as functions of concurrent tasks and recall type are shown in Table 4.3.

Table 4.3 Proportion of order errors in Experiment 4

	Baseline	Articulatory suppression	Tapping
Verbal recall	0.06 (0.06)	0.10 (0.07)	0.11 (0.17)
Action recall	0.05 (0.09)	0.06 (0.11)	0.08 (0.12)

A 3 × 2 (Concurrent task × Recall type) ANOVA showed no significant main effect of concurrent task, F (1.34, 29.39) = 1.796, $p = 0.190$, $\eta_p^2 = 0.075$, MSE = 0.010, no significant difference between verbal and action recall, F (1, 22) = 0.589, $p = 0.451$, η_p

$^2 = 0.026$, MSE = 0.007, and no significant interaction between concurrent task and recall type, F (1.34, 29.39) = 0.203, p = 0.726, $\eta_p{}^2 = 0.009$, MSE = 0.010.

Strategy report

Among the twenty-four participants, twenty-one participants reported using strategies. The scoring method was same as that in Experiment 3. The count scores and percentages of strategies as functions of recall type are shown in Table 4.4.

Table 4.4 Self-report strategies in Experiment 4

	Verbal recall (N=11)		Action recall (N=10)		Total (N=21)	
	Count	Percent	Count	Percent	Count	Percent
Visual tracking	6	55%	5	50%	11	52%
Imagining carrying out the actions	6	55%	4	40%	10	48%
Verbal rehearsal	2	18%	2	20%	4	19%
Decreasing interference	1	9%	0	0%	1	5%
Grouping actions	1	9%	0	0%	1	5%
Acronym	1	9%	1	10%	2	10%

Discussion

This experiment investigated the involvement of the visuospatial sketchpad in following spoken instructions, using the spatial tapping task as the concurrent interference task. The interference effect of the tapping task was relatively smaller than expected, failing to reach the conventional significance level of 5% (p = 0.073). One possibility is that the tapping task was relatively easy and became automatic after repetitive tapping. As evidence has shown that the activation of the brain area for action sequence planning (the supplementary motor area) decreases throughout action acquisition (Jenkins, Brooks, Nixon, Frackowiak, & Passingham, 1994; Seitz & Roland, 1992), it is possible that repetitive tapping did not place a sufficient load on the spatial working memory. Another reason for the absence of a significant effect is that the tapping keyboard was placed at the dominant-hand side of a participant, and was thus restricted to a small proportion of space relative to the whole display of objects; therefore tapping in such a limited space could hardly produce large interference to the spatial coding of the display. Moreover, the visual and spatial information of objects was still available during encoding, which allowed participants to continue using the visuospatial cues. This is supported by the strategy report, which indicated that 52 percent of responders eye-tracked the objects in space when the names of these objects were mentioned. Taken together, the spatial tapping task in this experiment was relatively weak in its exertion of interference to the visuospatial coding during the construction of representations of actions.

The articulatory suppression effect was evident in this experiment, consistent with previous findings (Allen, 2009; Allen & Gathercole, 2008), suggesting that the rehearsal mechanism was employed in remembering spoken instructions in a 3D instructional task. However, the articulatory suppression effect was no greater than the tapping effect, implying a similar degree of reliance on the two storage components in the working memory. In other words, both phonological coding and visuospatial coding were used in this 3D instructional task. Nevertheless, it is worth noting that articulatory suppression had a larger effect on memory of movement than tapping, whereas the reverse pattern was observed in memory of colour and object. These results suggest that the phonological loop and the visuospatial sketchpad may have different focuses in encoding different types of elements. The visual codes can be retained as locations in external space whereas the information of movement has to be rehearsed before the action can be imagined or executed.

The action advantage obtained in the previous experiment (Experiment 3) was replicated in this experiment. The contrasting pattern of serial position curves between the verbal and action recall conditions was also similar to that in Experiment 3, showing as a steep slope in the verbal recall condition versus a relatively flat line in the action recall condition (see curves of baseline conditions in Figure 4.2).

However, working memory did not interact with the recall type. Specifically, the similar spatial tapping effect in verbal and action recall was contrary to the expectation that there would be a greater contribution of the visuospatial sketchpad to the multimodal representation for actions than to the verbal representation for oral repetition. The action advantage still existed in the tapping condition. This might be due to the weak disruption effect of the tapping task and the accessibility of the visual display, which would mean that forming a multimodal representation of actions using this display was still feasible in the tapping condition. The similar articulatory suppression effect in verbal and action recall implies that there was a similar contribution of the phonological loop in forming representations for actions and for oral repetitions. This result is consistent with the findings of the last experiment, which suggested a basic supporting role of the phonological loop in maintaining spoken commands.

Nevertheless, it should be noted that the contribution from the phonological loop to the representations for the verbal and action recall differed in the different stages of encoding (see Figure 4.2). In the verbal recall conditions, the disruptive effect of articulatory suppression decreased across positions, implying a decreased contribution from the phonological loop during encoding. The action recall conditions, however, the disruptive effect of articulatory suppression was constant across all positions, implying a consistent reliance on the phonological loop throughout the encoding stage.

Similarly, the tapping effect also varied with the encoding process, which also differed in verbal and action recall conditions. In preparation for recall by actions, tapping was more disruptive to the memory of those actions in the middle of a sequence than to the first or last actions. First, this result indicates an increasing spatial demand during the beginning of encoding, which might have been due to the cumulative spatial memory load as a result of an increment in memory of locations for future actions. Second, the decreased tapping effect towards the end of sequence may be the result of the decrement of interference from tapping, which may reflect decreased attention for the control of tapping once it started. When an oral repetition was expected, the disruptive tapping effect decreased gradually during the encoding process, mainly reflecting the decreasing demand of attentional control over tapping. The lack of increment of the tapping effect in the earlier serial positions reflected little accumulation in the spatial load, suggesting that little spatial information was used in the beginning of encoding when an oral repetition was expected.

Although there are many differences in the way working memory contributes to verbal and action recall, these do not conflict with the assumption of a similar multimodal representation for both types of recall; rather, these differences imply that the process of forming a representation may differ depending on the purpose of recall. Moreover, these differences provide support for the multicomponent view of working memory, which assumes different innate processes for the different working memory components (Baddeley, 2000); they provide less support, however, to a unitary view that emphasizes the similarity between the memory activation processes among different modalities (Cowan, 1999).

As can be seen, the cognitive processes underlying the encoding of a sequence of actions in a rich environment is complicated; it is therefore necessary to simplify the representation formed to allow a better understanding of the process. One way of doing this is to block the entire visual display by asking participants to close their eyes during encoding. This would prevent the tracking of locations of objects and also block the encoding of the visual features of objects. If visual display of objects were indeed crucial in helping the formation of a multimodal representation in the 3D instructional task, taking away this external memory aid should disrupt this visuospatial encoding, and subsequently impair the recall. Moreover, the removal of visual display may force participants to rely on rehearsal, leading to a verbal-based representation for both types of recall. If this were true, the performances of two types of recall should be similar, and the action advantage should be partly or even totally eliminated. These hypotheses were tested in the next experiment.

Experiment 5

Introduction

This experiment set out to test the effect of removing the opportunity of encoding rich visual and spatial cues on the performance of recall of spoken instructions. This was achieved by requiring participants to close their eyes when listening to the instructions. It was predicted that eye closure would impair the performance of recall significantly.

Moreover, this experiment aimed to explore the contributions of visual support to the action advantage. In the discussion of the previous experiment, it was inferred that eye-closure should lead to the formation of a verbal-based representation for both types of recall, and thus would eliminate or decrease the action advantage. However, previous studies using a similar technique showed inconsistent results. In Allen's instructional task, participants were required to look away from the visual display when listening to the instructions. In one study, the absence of a visual display led to poor performance of recall and removed the action advantage (Allen, et al., 2009).

However, a later experiment failed to replicate the effect of absent visual display on action advantage; on this occasion, a lack of visual display impaired both types of recall similarly (Allen, 2011). Therefore, no specific hypothesis regarding the contribution of visual support to action advantage was made.

Method

Participants

Twelve native English speakers at the University of York were recruited through the electronic booking system in exchange for course credit or an honorarium of £4. None had taken part in the previous four experiments. There were 9 females and 3 males, aged from 18 to 47, with a mean age of 23.00.

Materials

The first three instruction lists in this experiment were the same as those in Experiment 4, and an additional list was added (see List 4 in Appendix 5). The setting of the visual display was the same as that in Experiment 4 except that there was no keypad for tapping.

Design

In a 2 $\times$ 2 within-subject design, the two independent variables were eye-closure and recall type. Participants either opened or closed their eyes during encoding, followed by either verbal or action recall. The measurements made were the same as those in Experiment 4.

Procedure

Participants signed the consent forms and were introduced to the task, followed by a practice round with two practice trials for each condition. All participants finished four conditions. The eye-open conditions were the same as the baseline conditions in Experiment 4. In the eye-closure conditions, participants were instructed to close their eyes throughout the presentation of instructions until the beep sound, upon which they opened their eyes and began recall. As in the previous experiments, the verbal recall required repetition of instructional sentences, whereas the action recall required manipulations of objects by hand. In both types of recall, the importance of sequence was emphasized. At the end of each condition, participants were asked about the strategies they employed in the experiment.

Results

Actions

The serial recall of actions was scored by averaging the correct actions per instruction across 12 formal trials. The scoring method was same as that in previous experiments. The means and the standard deviations of actions as functions of eye closure and type of recall are displayed in Table 4.5.

Table 4.5 Means (and standard deviations) of actions in Experiment 5

	Verbal recall	Action recall	Means
Eyes closure	1.50 (0.59)	2.01 (0.50)	1.76 (0.50)
Eyes open	2.80 (0.74)	3.67 (0.70)	3.24 (0.67)
Means	2.15 (0.57)	2.84 (0.56)	2.50 (0.53)

A 2 × 2 (Eye-closure × Recall type) ANOVA revealed significant main effect of eye-closure, F (1, 11) = 98.398, p < 0.001, $\eta_p{}^2$ = 0.899, MSE = 0.266, significant main effect of recall type, action recall was better than verbal recall, F (1, 11) = 39.744, $p < 0.001$, $\eta_p{}^2 = 0.783$, MSE = 0.144, and no significant interaction between eye-closure and recall type, F (1, 11) = 3.430, $p = 0.091$, $\eta_p{}^2 = 0.238$, MSE = 0.110.

Dependent *t*-tests with Bonferroni corrections showed significant eye-closure effect in both verbal and action recall conditions (*ps* < 0.01). Independent *t*-tests with Bonferroni corrections showed that action advantage existed in the eye-open condition as well as in the eye-closure condition (*ps* < 0.05).

Elements

In this experiment, each action contained three elements: movements, colours, and objects. The scoring method was same as that in previous experiments. As in the previous experiments, the chance levels for elements were different (see Experiment 3 for details).The means and standard deviations of the elements and colourful objects as functions of eye closure and type of recall are presented in Table 4.6.

Table 4.6 Means (and standard deviations) of elements in Experiment 5

		Verbal recall	Action recall	Means
Movement	Eye closure	2.39 (0.64)	3.03 (0.46)	2.71 (0.46)
	Eye open	3.27 (0.70)	3.96 (0.65)	3.62 (0.63)
	Means	2.83 (0.64)	3.49 (0.51)	3.16 (0.52)
Colour	Eye closure	2.13 (0.68)	2.73 (0.60)	2.43 (0.60)
	Eye open	3.70 (0.62)	4.20 (0.55)	3.95 (0.52)
	Means	2.91(0.53)	3.47 (0.51)	3.19 (0.49)
Object	Eye closure	2.34 (0.81)	3.10 (0.55)	2.72 (0.61)
	Eye open	3.74 (0.55)	4.31 (0.50)	4.03 (0.49)
	Means	3.04 (0.56)	3.70 (0.48)	3.37 (0.49)
Colourful Object	Eye closure	1.93 (0.74)	2.65 (0.60)	2.30 (0.62)
	Eye open	3.60 (0.64)	4.20 (0.57)	3.90 (0.55)
	Means	2.77 (0.54)	3.43 (0.52)	3.09 (0.51)

A 2 × 2 × 3 ANOVA (Eye-closure × Recall type × Element) was conducted. There was significant main effect of eye-closure, F (1, 11) = 121.977, $p < 0.001$, $\eta_p^2 = 0.917$, MSE = 0.459, recall type, F (1, 11) = 38.591, $p < 0.001$, $\eta_p^2 = 0.778$, MSE = 0.366, and element, F (1.16, 12.71) = 7.974, $p = 0.012$, $\eta_p^2 = 0.420$, MSE = 0.134. Element significantly interacted with eye-closure, F (1.09, 11.94) = 10.370, $p = 0.007$, $\eta_p^2 = 0.485$, MSE = 0.208, but there was no other two-way or three-way interaction ($ps > 0.05$).

One-tailed dependent t-tests with Bonferroni corrections showed that eye-closure effect was significant in all elements and colourful objects ($ps < 0.05$); and action advantage also existed in all elements and colourful objects ($ps < 0.05$).

Planned contrasts showed larger eye-closure in colour than in object, F (1, 11) = 28.813, $p < 0.001$, $\eta_p^2 = 0.724$, MSE = 0.019, which in turn larger than in movement, F (1, 11) = 6.088, $p = 0.031$, $\eta_p^2 = 0.356$, MSE = 0.319.

Binding

The role played by the visual display in binding was examined using the same method as in previous experiments. If the visual display has a role in binding elements in an action, eye closure should have a larger disruptive effect on memory of bound entities than on individual elements.

The effect of eye-closure on binding colour to object was tested using a 2 × 2 × 2 (Eye-closure × Binding × Recall type) ANOVA, the variable binding included colourful object and object, and the variable eye-closure included the eye-open and eye-closure conditions. Results showed significant main effect of eye-closure, F (1, 11) = 86.201, $p < 0.001$, $\eta_p^2 = 0.887$, MSE = 0.591, binding, F (1, 11) = 131.842, $p < 0.001$, $\eta_p^2 = 0.923$, MSE = 0.014, and recall, F (1, 11) = 44.934, $p < 0.001$, $\eta_p^2 = 0.803$, MSE = 0.232. As predicted, the interaction between eye-closure and binding was significant, with a larger eye-closure disruptive effect colourful object than on object, F (1, 11) = 31.847, $p < 0.001$, $\eta_p^2 = 0.743$, MSE = 0.016. There was no other two-way or three-way interaction ($ps > 0.05$).

The effect of eye-closure on binding movement to corresponding colourful object was tested using a 2 × 2 × 2 (Eye-closure × Binding × Recall type) ANOVA, the variable binding included colourful object and action. Results showed significant main effect of eye-closure, F (1, 11) = 108.655, $p < 0.001$, $\eta_p^2 = 0.908$, MSE = 0.524, binding, F (1, 11) = 175.296, $p < 0.001$, $\eta_p^2 = 0.941$, MSE = 0.049, and significant recall type, F (1, 11) = 45.550, $p < 0.001$, $\eta_p^2 = 0.805$, MSE = 0.238. There was no significant interaction between eye-closure and binding, F (1, 11) = 1.069, $p = 0.323$, $\eta_p^2 = 0.089$, MSE = 0.091, nor there was any other two-way interaction ($ps > 0.05$). There was a significant three-way interaction, F (1, 11) = 32.335, $p < 0.001$, $\eta_p^2 =$ 0.746, MSE = 0.011. The interaction between eye-closure and binding was examined separately in verbal and action recall conditions, and when oral repetition was required, the eye-closure was larger in colourful object than in action, F (1, 11) = 6.509, $p = 0.027$, $\eta_p^2 = 0.372$, MSE = 0.062, however, when action recall was required, there was no interaction between eye-closure and binding, F (1, 11) = 0.965, $p = 0.347$, $\eta_p^2 = 0.081$, MSE = 0.039.

Serial positions

Each position was coded for the percentage of correct trials, ranged from 0 to 1. A 2 × 2 × 5 (Eye-closure × Recall type × Serial position) ANOVA found significant main effect of eye-closure, F (1, 11) = 91.580, $p < 0.001$, $\eta_p^2 = 0.893$, MSE = 8.491, recall type, F (1, 11) = 29.803, $p < 0.001$, $\eta_p^2 = 0.730$, MSE = 5.818, and serial position, F (4, 44) = 33.232, $p < 0.001$, $\eta_p^2 = 0.751$, MSE = 3.599. Serial position showed significant interaction with eye-closure, F (4, 44) = 3.133, $p = 0.024$, $\eta_p^2 = 0.222$, MSE = 2.768, and also with recall type, F (4, 44) = 5.837, $p < 0.001$, $\eta_p^{\,2}$ =0.347, MSE = 2.065. There was no other two-way or three-way interaction (*ps* > 0.05).

The post-hoc tests with Bonferroni corrections indicated that decrement of recall was significant between positions 1 and 2 (p = 0.021), and positions 2 and 3 (p < 0.001), but not between the other adjacent positions (*ps* > 0.05). Simple effect analyses showed that action advantage existed in all positions (*ps* < 0.001), and eye closure effect existed in all positions (*ps* < 0.001).

Planned contrasts were conducted to locate the interaction between eye-closure effect and serial position, which occurred in the last two positions, with greater eye-closure effect in position 4 than in position 5 (p = 0.016). The interaction between recall type and serial position also occurred in the last two positions, with a larger action advantage at position 5 than at position 4 (p = 0.010). These effects are presented in Figure 4.3.

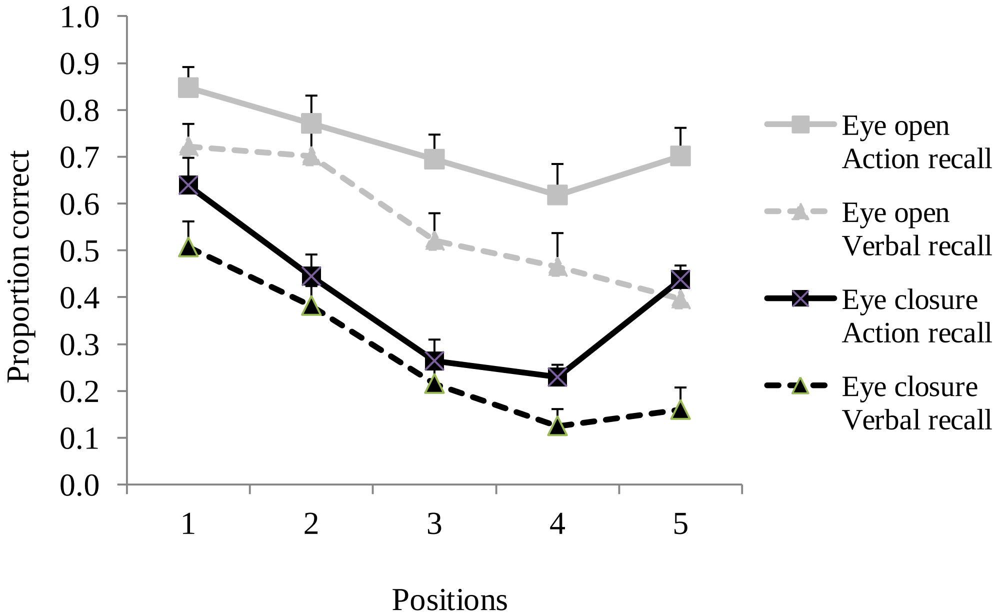

Figure 4.3 Serial position curves (means and standard errors) as functions of eye-closure and type of recall in Experiment 5.

Proportions of order errors

The way of calculating proportions of order errors was same as that in Experiment 3. The means and standard deviations of performance of recall as functions of eye-closure effect and the recall type are presented in Table 4.7.

Table 4.7 Proportion of order errors (and standard deviations) in Experiment 5

	Verbal recall	Action recall
Eye closure	0.18 (0.16)	0.12 (0.14)
Eye open	0.06 (0.06)	0.06 (0.07)

A 2 × 2 (Eye-closure × Recall type) ANOVA was conducted. There was a significant main effect of eye-closure, F (1, 11) = 7.121, p = 0.022, η_p^2 = 0.393, MSE = 0.014, but no significant main effect of recall type, F (1, 11) = 1.283, p = 0.281, η_p^2 = 0.104, MSE = 0.009, and no significant interaction between eye-closure and recall type, F (1, 11) = 3.137, p = 0.104, η_p^2 = 0.222, MSE = 0.002. One-tailed dependent t-tests with Bonferroni correction showed that the eye-closure effect was significant in the verbal recall condition, t (11) = 2.926, p = 0.014, but not in the action recall condition, t (11) = 1.950, p = 0.078.

Strategy report

All twelve participants indicated using multiple strategies in the 3D instructional task. The count score stands for the number of responders that reported using that strategy. The percentage score stands for the percentage of responders used that strategy. These scores as functions of eye closure and type of recall are shown in Table 4.8.

Table 4.8 Self-report strategies in Experiment 5

Verbal recall (N=12)	Eye open		Eye closure		Subtotal	
	Count	Percent	Count	Percent	Count	Percent
Imagining carrying out the actions	3	25%	5	42%	8	33%
Visual tracking	7	58%	0	0%	7	29%
Verbal rehearsal	0	0%	5	42%	5	21%
Grouping actions	0	0%	1	8%	1	4%
Action recall (N=12)	**Eye open**		**Eye closure**		**Subtotal**	
	Count	Percent	Count	Percent	Count	Percent
Imagining carrying out the actions	3	25%	6	50%	9	38%
Visual tracking	8	67%	0	0%	9	38%
Verbal rehearsal	0	0%	4	33%	4	17%
Grouping actions	0	0%	0	0%	0	0%
Total (N=12)	**Eye open**		**Eye closure**		**Total**	
	Count	Percent	Count	Percent	Count	Percent
Imagining carrying out the actions	6	25%	11	46%	17	35%
Visual tracking	15	63%	0	0%	15	31%
Verbal rehearsal	0	0%	9	38%	9	19%
Grouping actions	0	0%	1	4%	1	2%

Discussion

This experiment set out to test the effect of eye closure on remembering instructions in the 3D instructional task. Results showed that blocking the encoding of visual and spatial information by closing the eyes led to a significant impairment of recall, consistent with the first hypothesis and also existing literature (Allen, 2009). However, there was no interaction between eye-closure and recall type. These results suggest that the visual support benefited the process of memorizing a series of action commands upon objects in a wide space, and this benefit was similar to the formation of representations for verbal and action recall.

The benefit of visual display manifested in several aspects. A major role of the visual display was found to be related to the maintenance of visuomotor information, supported by the finding of a significant disruptive effect of eye-closure on memory of movement, colour and object. The greater disruption of eye closure on memory of colour and object than movement also suggests a greater contribution of the visual display to retaining visual features compared to motor information. In fact, the absence of the visual display reversed the pattern of a superior recall of colour to movement; that is, eye-closure led to poorer recall of colour than recall of movement. This result suggests the superior memory of colour to movement was mainly attributed to the support of the visual display.

Besides maintaining visual features in an action, visual support also helped bind these elements, with evidence from the greater eye-closure effect on the memory of colourful objects than on the memory of objects alone. This suggests the visual support has a role in maintaining the colour and object as a bound entity, which is consistent with the notion that visual short-term memory plays a role in binding visual features (Brockmole, 2009). On the other hand, the similar impairment by eye-closure on colourful objects and actions implies the visual support did not have help link movement with the associated object.

The third role of the visual display is related to sequencing. This was supported by the increased proportion of order errors occurring in the eye-closure condition, suggesting that the visual support might have a role in maintaining the sequencing related to actions. Exactly how visual support contributed to the sequencing can be discerned from the report of strategies used, in which 63 percent of participants indicated that they were visually tracking the objects in space while listening to the spoken commands. Thus, visual support might support the encoding and maintenance of the ordinal information in an instruction by helping the participants to mentally draw a route representing the sequence of operations upon to-be-enacted objects in space. In this sense, visual support served as an external memory that helped offload the burden of remembering multiple visually-rich objects in space, consistent with the notion of deictic pointers in space (Spivey, et al., 2004). Moreover, the process of active tracking of the locations of objects is similar to the inner scribe in the visuospatial sketchpad described by Logie, which has the function of retaining a sequence of movements around an array of locations (Logie, 2011). Furthermore, the eye-closure effect was found to be larger in the earlier positions than in the last position (also see Figure 4.3), implying a greater contribution from the visual support to the memory of actions earlier in the sequence. This greater reliance on the visual support may reflect the active tracking of the locations of to-be-performed objects in the beginning of the playback of an instruction; however, this tracking may become hard to keep up with the procession of instructions, and therefore towards the end of instructions verbal rehearsal may instead be used.

Self-reported strategies during encoding changed as a consequence of eye closure. Visual tracking was preferred when coding the locations of objects was feasible, but when it became impossible in the eye-closure conditions, participants shifted to use strategies like rehearsal and 'imagining carrying out the action ' more often. This means that in an instructional task with rich visual cues, the visual support was probably a more efficient and convenient way of coding and maintaining the information than the other strategies, and therefore participants took advantage of the visual support as long as they had the chance to do so. The flexible use of strategies also corresponds to the viewpoint that, when completing a cognitive demanding task, participants can flexibly adopt strategies depending on the situation (Logie, 2011).

This experiment again replicated the findings of action advantage, consistent with findings in Experiments 3 and 4, and also existing literature (Allen, 2009;

Gathercole, et al., 2008). Interestingly, the serial position curves were found to be different in the verbal and action recall conditions. Specifically, the action advantage was larger in the last than in the earlier positions; in other words, the recency effect was present in the action recall conditions but not in the verbal recall condition (see Figure 4.3). Several factors are known to contribute to the rise of recency effect, such as greater temporal distinctiveness, less retroactive interference, and, most importantly, more response suppressions (also see the section of sequential representations in Chapter 1). The suppression of earlier items tends to make room for the later items, and therefore the last item has the fewest candidate items in memory (Farrell & Lewandowsky, 2002). The evident recency effect in the action but not in the verbal recall condition thus implies more suppression of performed than orally-reported actions. It can be speculated that this effective suppression of enactment may be due to the visible completion status of these enactments. For example, the presence of a red pencil in the black box may serve as a reminder for the completion of that action, thus preventing potential erroneous actions on the same red pencil, and also leaves fewer candidate actions for the future. In contrast, oral repetitions are unlikely to leave lasting traces of completion; therefore the suppression of the actions that have been recalled was weak, and this consequently increases the chance of repetition errors. Moreover, the sound traces of one's own oral repetitions may interfere with the existing verbal representations of to-be-recalled actions, deteriorating the later items.

In this experiment, there was no significant interaction between eye-closure and recall type. It was originally inferred that the lack of visual support should led to a verbal-based representation of instructions for both verbal and action recall, resulting in similar performances of recall. One fallacy of this conjecture was that the representations in eye-closure conditions were not pure verbal-based representations. In the eye closure conditions, 46 percent of participants imagined themselves doing the actions and 38 percent of participants rehearsed the spoken commands, suggesting that the representations in eye-closure conditions were mixed rather than purely verbal-based. Moreover, both recall groups adopted the rehearsal and ‘imagining carrying out action’ strategy, suggesting that both recall groups might have formed similar representations. Thus, the remaining action advantage in the eye-closure conditions probably reflects the benefit of performance of oral repetition in the output stage. It is also worth mentioning that in other aspects of measurements, such as elements, binding colour and object, the eye-closure effects were similar in the two recall groups, reinforcing the theory that visual support plays a similar role in representing instructions for the two types of recall.

Nevertheless, a trend of reduction of the action advantage by eye closure was still observable, reflected in the large effect size of interaction between eye closure and recall of actions (partial eta square equals 0.238). Given that there were only twelve participants, the power for detecting the different contributions of visual support to verbal and action recall is relatively low. Therefore, it is possible that visual support may have contributed somewhat to the action advantage, but that this contribution was too small to reach the conventional level of significance.

Taken together, visual support was found to play an important role in maintaining and binding visual aspects of information in an action as well as keeping sequence of actions in memory. Moreover, these roles of visual support in the formation of representations for verbal and action recall were found to be similar.

General discussion

This chapter investigated the role of the visuospatial sketchpad in supporting the encoding of spoken instructions. In Experiment 4, it has been found that interfering with spatial coding by the spatial tapping task led to a relatively small decrement in recall performance. In Experiment 5, when visuospatial coding was completely blocked by eye-closure, the performance of both verbal and action recall dropped significantly. The results of the two experiments thus suggest the importance of coding visual and spatial information in a rich task environment even when instructions are delivered in a verbal-based form. The benefits of including the visuospatial codes in forming a multimodal representation are likely to be reflected in a variety of factors, including better maintenance of elements in an action, binding of visual features like colour to an object, and also the retaining of ordinal information.

In the two experiments, the finding of a superior recall by actions to oral repetition was replicated, thus establishing the action advantage in the 3D instructional task. However, in both experiments, there was no interaction between the visuospatial sketchpad and recall, suggesting that the benefit of action advantage is neither from spatial coding nor from the visual support.

In summary, the greatest contribution of working memory to instruction-following in this task appears to be the central executive, which has been shown to contribute substantially to the maintenance and binding of visual features as well as to the retaining of ordinal information of actions. Importantly, all these functions are likely to relate to the attentional control of eye movement in forming a spatial representation of future actions. The contributions of the visuospatial sketchpad have been summarized above; one thing worth noticing is the roles that overlap with the central executive, including binding visual features and retaining sequence of actions. Finally, the phonological loop supports the encoding of all types of elements in an action, with a focus on remembering certain type of elements that are difficult to maintain in other ways, such as the motor information in an action. Therefore, all three working memory components are involved in supporting following instructions in a 3D environment, which is not uncommon in a complex task (Logie, 2011).

The absence of a disruptive effect of any of the concurrent tasks on the action advantage in following instructions in this task suggests that the benefit of action advantage cannot be attributed to the functioning of working memory during encoding. The strategy report showed that participants in the two recall groups used similar strategies, therefore excluding strategies as the source of action advantage. However, adopting similar strategies in two types of recall did not guarantee the equal quality of the representation formed. It is possible that there were benefits beyond the conscious awareness of participants during the process of constructing representations for actions. It is also possible that cognitive functions other than working memory might have contributed or were the main source of the action advantage.

The following chapter reports two experiments designed to investigate the roles played by working memory in following instructions that were written rather than spoken.

Chapter 5

The role of working memory in following written instructions

Introduction

Previous experiments indicated significant involvement of working memory in following spoken instructions and have established the action advantage in the 3D instructional task. This chapter aims to extend these findings to written instructions. Two experiments were carried out to examine the contributions of working memory to encoding written instructions using the dual task paradigm. Specifically, three subcomponents under the multicomponent working memory model were investigated, namely, the phonological loop, the visuospatial sketchpad and the central executive (Baddeley & Hitch, 1974). The 3D instructional task was used with instructions shown as texts on a computer screen. Another aim of these two experiments was to test whether the action advantage, the phenomenon of a superior performance of recall by actions than by oral repetitions, also exists in written instructions. They also aimed to test whether working memory helps explain this benefit.

First, the phonological loop should be considered, which has been found to be involved in encoding spoken instructions. Unlike spoken instructions that can gain direct access to the phonological store (Baddeley, et al., 1984), visually presented information has to be transformed to phonological recoding through the orthographic. This grapheme-to-phoneme conversion recodes written materials into phonological representations (Conrad, 1964; Vallar & Papagno, 2002), and this recoding relies on the rehearsal mechanism in the phonological loop (Baddeley & Larsen, 2007). Another contribution of the phonological loop is related to the cognitive process of reading. Eye movement research shows that the phonological loop is involved in integrating information across saccades during reading (Rayner, 1998). Therefore, the phonological loop is expected to contribute significantly to following written instructions. As a result, articulatory suppression should disrupt the recoding process as well as the rehearsal process, and consequently impair the recall.

Secondly, the visuospatial sketchpad should be considered. Previous experiments in this study indicated that the visuospatial sketchpad was involved in retaining and binding the visual features of an object, and also had a role in maintaining sequences of actions. These roles are all related to the rich environment of the 3D instructional task, and should therefore also be similar in following written instructions using the same 3D instructional task. In situations involving following written instructions, a unique role of the visuospatial sketchpad is to maintain the visual forms of words. Therefore, the visuospatial sketchpad is expected to contribute to the process of remembering written instructions; the task that interferes with the visuospatial sketchpad should thus impair the recall of written instructions.

It is not uncommon for visual codes to be used in coding verbal material in the immediate verbal serial recall (Logie, 2003; Logie, et al., 2000; Posner, Boies, Eichelman, & Taylor, 1969; Posner & Keele, 1967). Moreover, these visual codes are used even when phonological codes are still available (Logie, 2003, 2011). It thus appears that the coding of written instructions is complicated; it is therefore worth paying heed to the strategies employed by participants. In a study specifically addressing this issue, large variations of individual strategies were found (Logie, Della Sala, Laiacona, Chalmers, & Wynn, 1996). Besides common strategies like rehearsal and visual coding of the word forms, participants also used strategies such as semantic coding, first-letter acronyms, chunking, and visual imagery to encode the verbal material, and many participants used more than one strategy. Both word length effect and phonological similarity effect were affected by the strategy adopted. Therefore, it is important to know the strategies adopted by participants in a verbal task. Thus, in the experiments described in this chapter, participants were asked about their strategies and also the way in which they read the instructions.

The central executive has been found to have played a substantial role in following spoken instructions, including the attentional control of eye tracking of objects in building a spatial representation of actions and binding visual features of an object. With regards to following written instructions in the 3D task environment, dividing and shifting attention is important. For example, as the instructions were lengthy and beyond the capacity of the phonological loop, participants might have had to use the locations as temporary caches, and thus they would have looked at the objects in display from time to time. This frequent switch between looking at written instructions and referring to the objects in display can be attention-demanding, and place a heavy load on the central executive. Therefore, the central executive is expected to be highly involved in encoding written instructions. The same interference task, the backward counting task, was used to disrupt the central executive. Similar as that in the previous experiments, the contribution of the phonological loop to retaining the intermediate products in the backward counting task had to be controlled. Therefore, the additional deduction of the backward counting task from the articulatory suppression task represents the involvement of the central executive, known as the backward counting effect.

The second aim of the experiments presented in this chapter is to extend the phenomenon of action advantage, the benefit of recalling by actions compared to oral repetitions of written instructions. Literature on the memory of to-be-enacted actions has already demonstrated the existence of the action advantage in following written instructions (Koriat, et al., 1990). Importantly, the results of Koriat et al.'s study showed that the action advantage arose from the encoding stage rather than the retrieval stage. This is because the intention of forming an action-based representation that stresses future enactment took advantage of imaginal-enactive properties of envisaged acts, which is superior to the abstract proposition form of a verbal representation used when oral repetition was required. Given this literature and the evident action advantage in following spoken instructions, the superior recall by actions than by oral repetition would be expected in following written instructions.

Working memory has not been found to interact with recall type in any experiments exploring the relations between working memory and recall type in following spoken instructions (Allen, 2009; also Experiment 3, 4 and 5 in this study). Both the phonological loop and the central executive have been shown to make similar contributions to the two types of recall, as has the visuospatial sketchpad, which has occasionally had a tendency to make a greater contribution to action recall. Therefore, no specific hypothesis regarding the interaction between working memory and types of recall was made for these two experiments.

Experiment 6 investigated the roles played by the phonological loop and central executive in supporting the encoding of written instructions using the articulatory suppression and backward counting task respectively. Experiment 7 investigated the involvement of the visuospatial sketchpad in encoding written instructions using a spatial tapping task as the concurrent interference task. In both experiments, participants were divided into two groups: one group verbally repeated the instructions and the other group performed the actions.

Experiment 6

This experiment investigated the contributions of the phonological loop and the central executive to the process of encoding written instructions using the same interference tasks as Experiment 3, i.e. the articulatory suppression task and the backward counting task. The articulatory suppression task required participants to repeat irrelevant digits continuously during the time that they were reading the instructions. The backward counting task required continuous subtraction of two from a three-digit number during reading.

There were three conditions in this experiment: the baseline condition of no concurrent task, articulatory suppression, and backward counting. The baseline condition served as a comparison condition for the articulatory suppression condition; the difference in performance of the two conditions is the articulatory suppression effect, representing the involvement of the phonological loop. The difference in performance between the backward counting condition and articulatory suppression condition is the backward counting effect; this represents the central executive.

Both the phonological loop and central executive were expected to play important parts in following written instructions, and therefore the articulatory suppression and backward counting effects were expected; this formed the first hypothesis. In the previous experiments using spoken instructions, the backward counting effect was significantly larger than the articulatory suppression, suggesting a greater contribution from the central executive compared to the phonological loop. The same trend was expected in written instructions; the second hypothesis, therefore, was a larger backward counting effect compared to the articulatory suppression effect in following written instructions. Finally, the third hypothesis related to the effect of recall. Specifically, the performances of recall by actions were compared to performances of repeating the instructions verbally, and an action advantage was anticipated.

Careful consideration was given to the particular format of the written instructions. Research has shown that syntax affects the ability to remember sentences and execute commands; for example, sentences using main clauses require less time than sentences with implicit embedded clause (Seymour, 1974; Wright & Wilcox, 1978). Therefore the written instructions in this experiment were kept simple, as action phrases using an imperative sentence structure, and conjunctive words like ‘then’ and ‘and’ were not included.

There are several options of presenting written instructions. One way is the rapid serial visual presentation (Masson, 1983), in which words or action phrases are presented at a set rate in the same spatial location. This method provides most resemblance to spoken instructions in the way that, once displayed, the instructions disappear unless rehearsed. However, given the large individual differences in reading and comprehension, it is difficult to set an appropriate duration of the presentation of each word or action. Moreover, the time spent viewing a word is related to the conceptual processing of the word, and therefore a prematurely determined viewing time will impair the comprehension (Just & Carpenter, 1980), consequently impairing the action planning during the course of reading written instructions. Therefore, rapid serial visual presentation was not appropriate for this experiment. Self-paced presentation, on the other hand, allows each action phrase to appear one at a time, and readers advance the text by pushing a button (Rayner, 1998). However, this method may require additional motor movement. Therefore, the best method of presentation is also the one that most resembles people’s daily experience of reading instructions, that is, to present all steps of actions simultaneously. This method also allows readers more freedom to read them according to their own pace and habits. Therefore, in this experiment, all action phrases in an instruction message would be shown simultaneously on the computer screen and disappear at the same time.

Method

Participants

Twenty-four native English speakers at the University of York were recruited through the electronic booking system in exchange for course credit or an honorarium of £6. None had taken part in the previous five experiments. There were 20 females and 4 males, aged from 18 to 28, with a mean age of 19.83.

Materials

The instructions were same as those in Experiment 4. There were four types of action phrases (touch, pick up…put it into, push and spin) and twelve items of colourful stationery, including six small objects (a yellow ruler, a blue ruler, a white eraser, a green eraser, a red pencil and a black pencil) and six containers (a black box, a red box, a yellow bag, a white bag, a blue folder and a green folder). Each instruction contained five action phrases with no repetition of objects, and adjacent objects were always in different colours. A typical instruction was ‘push red box, pick up black pencil, put it into yellow bag, touch red pencil, spin blue ruler’, with each presented as a separate row of text. The action phrases in an instruction were presented simultaneously for 13 seconds, the same as the duration of a typical spoken instructional sentence in Experiments 4 and 5.

Three lists of instructions were adapted from those in Experiment 4 with the exclusion of conjunction words (see Appendix 6). Each list contained 14 instructional trials, twelve formal trials and two spare trials. In addition, three practices containing six trials, two for each condition were prepared. There were a total of 32 different three-digit numbers for the articulatory suppression and backward counting condition, same as those in Experiment 3 (see Appendix 4). Lists of instructions were programmed and played using the Eprime software.

All items of stationery were placed on a 146 cm (length) × 75 cm (width) × 71 cm (height) desk. The locations of the stationery remained the same within each individual trial but varied between trials. A monitor displaying written instructions sat behind the stationery (see Figure 5.1).

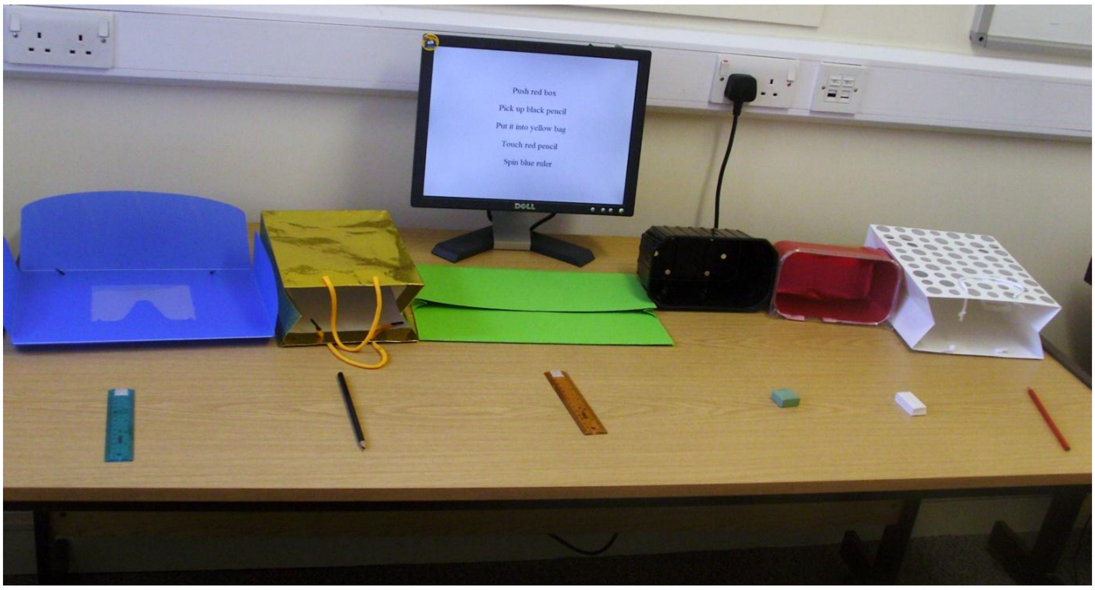

Figure 5.1 Display of following instruction task in Experiment 6.

To increase the number of participants who responded to questions on strategy, the strategy questionnaire in this experiment provided six options of strategies: rehearsal, remembering the words visually, Imagining doing it in head, grouping the actions, using acronyms, and using no strategies. The probe question used in the previous experiments was also provided, that is, ‘If you are using any other strategy, please state this’ (see Appendix 7).

Design

In a 3×2 mixed design, concurrent task was set as a within-subject variable, including baseline, articulatory suppression and backward counting conditions, and recall type was a between-subject variable, including verbal and action recall. The main dependent variable was the mean number of correct actions per instruction sequence. Other measurements included elements (movement, colour, and object), combined elements (colourful object), and also the percentage of correct trials in each serial position.

Procedure

The experiment was carried out in a quiet room. Upon arrival, each participant was randomly assigned to one of the recall groups. They were then introduced to the instruction task, and carried out the six-trial practice for all conditions. All participants sat at the desk, facing the monitor and the display of objects. The experimenter sat at another desk 30 cm away from the participants, controlling the delivery of instructions.

Each participant finished three conditions, baseline, articulatory suppression and backward counting condition. Three sets of instruction sequences were created, with each list containing 12 sequences. These sequence sets were implemented in counterbalanced order for each participant, balanced out across each concurrent task condition (with each condition utilizing the same sequence set an equal number of times over the study).

In all conditions, the entire instructional sequence (containing 5 action segments) was simultaneously presented on screen in *Times new roman* font, size 16, for 13s. Each of the 5 action segments appeared on a different line, aligned to the screen centre. This was followed by a one second blank screen delay and then a beep sound indicating recall. In the articulatory suppression conditions, participants first saw a three-digit number (e.g. *358*) in the centre of the screen (in the same font type and size as instructions) for 3s and began repeating it continuously (‘3’-‘5’-‘8’-‘3’-‘5’-‘8’…) at a paced speed of two seconds per cycle, throughout instruction presentation. The procedure of the backward counting conditions was similar, except that participants counted backwards from the three-digit number in decrements of two, for example, ‘3’-’5’-’8’- ‘3’-’5’-’6’- ‘3’-’5’-’4’... etc.

According to the assigned group, a participant either repeated the instructions back (verbal recall) or performed the actions (enactment recall), with the experimenter recording these responses. Oral repetition was recorded by the Audacity software and actions were videotaped. The experimenter also kept a written record of the recall. At the end of each trial, the experimenter changed the locations of objects randomly on the table while participants closed their eyes. The strategy questionnaire was given at the end of the experiment. Participants were also interviewed about the way they read the instructions.

Results

Actions

Serial recall of actions was scored by averaging the number of correct actions in each instructional sequence. The scoring method was same as that in previous experiments. Means and the standard deviation of actions as functions of concurrent task and type of recall were displayed in Table 5.1.

Table 5.1 Means (and standard deviations) of actions in Experiment 6

	Verbal recall	Action recall	Means
Baseline	3.49 (0.89)	4.18 (0.43)	3.84 (0.77)
Articulatory suppression	2.98 (0.64)	3.67 (0.59)	3.32 (0.70)
Backward counting	1.89 (0.70)	2.54 (0.43)	2.21 (0.66)
Means	2.79 (0.57)	3.46 (0.34)	3.12 (0.57)

A 3 × 2 (Concurrent task × Recall type) ANOVA showed significant main effect of concurrent task, F (2, 44) = 61.017, $p < 0.001$, $\eta_p^2 = 0.735$, MSE = 0.271, and significant main effect of recall type, action recall was better than verbal recall, F (1, 22) = 12.509, $p = 0.002$, $\eta_p^2 = 0.362$, MSE = 0.219. There was no significant interaction between concurrent task and recall type, F (2, 44) = 0.014, $p = 0.986$, $\eta_p^2 = 0.001$, MSE = 0.271.

Planned contrast indicated significant articulatory suppression effect, F (1, 22) = 11.511, $p = 0.003$, $\eta_p^2 = 0.344$, MSE = 0.547, and significant backward counting effect, F (1, 22) = 56.894, $p < 0.001$, $\eta_p^2 = 0.721$, MSE = 0.520. There was no interaction between articulatory suppression and recall type, F (1, 22) < 0.001, $p = 0.998$, $\eta_p^2 < 0.001$, MSE = 0.547, and no interaction between backward counting and recall type, F (1, 22) = 0.021, $p = 0.887$, $\eta_p^2 = 0.001$, MSE = 0.520.

Individual scores of the articulatory suppression effect and backward counting effect were calculated for each participant. The 2 × 2 (Effect × Recall type) ANOVA found significantly larger backward counting effect compared to the articulatory suppression effect, F (1, 22) = 5.448, $p = 0.029$, $\eta_p^2 = 0.198$, MSE = 0.788. There was no significant main effect of recall type, F (1, 22) = 0.020, $p = 0.888$, $\eta_p^2 = 0.001$, MSE = 0.139, and no interaction between effect and recall type, F (1, 22) = 0.007, $p = 0.936$, $\eta_p^2 < 0.001$, MSE = 0.788.

One-tailed dependent *t*-tests with Bonferroni corrections showed significant articulatory suppression effect in both recall groups (*ps* < 0.05), and significant backward counting effect in both recall groups (*ps* < 0.05). One-tailed independent *t*-tests with Bonferroni corrections found the presence of action advantage in baseline conditions as well as in the two concurrent task conditions (*ps* < 0.05).

Elements

Each action contained three elements: movement, colour, and object. The method of scoring was same as that in the previous experiments. As in the previous experiments, the chance levels for elements were different (see Experiment 3 for details).The means and standard deviations of the elements and colourful objects are presented in Table 5.2.

Table 5.2 Means (and standard deviations) of elements in Experiment 6

		Verbal recall	Action recall	Means
Movement	Baseline	3.83 (0.79)	4.43 (0.43)	4.13 (0.69)
	Articulatory suppression	3.30 (0.61)	3.87 (0.52)	3.58 (0.62)
	Backward counting	2.29 (0.69)	2.89 (0.50)	2.59 (0.67)
	Means	3.14 (0.52)	3.73 (0.34)	3.43 (0.53)
Colour	Baseline	4.04 (0.82)	4.49 (0.33)	4.26 (0.65)
	Articulatory suppression	3.65 (0.66)	4.22 (0.52)	3.93 (0.65)
	Backward counting	2.45 (0.63)	3.11 (0.47)	2.78 (0.64)
	Means	3.38 (0.60)	3.94 (0.30)	3.66 (0.55)
Object	Baseline	3.98 (0.82)	4.51 (0.33)	4.25 (0.67)
	Articulatory suppression	3.70 (0.65)	4.26 (0.50)	3.98 (0.64)
	Backward counting	2.49 (0.67)	3.16 (0.48)	2.83 (0.66)
	Means	3.39 (0.60)	3.98 (0.30)	3.69 (0.55)
Colourful object	Baseline	3.94 (0.85)	4.47 (0.34)	4.20 (0.69)
	Articulatory suppression	3.59 (0.68)	4.20 (0.52)	3.90 (0.67)
	Backward counting	2.37 (0.66)	3.04 (0.44)	2.70 (0.64)
	Means	3.30 (0.62)	3.90 (0.30)	3.60 (0.57)

A 3 × 2 × 3 (Concurrent task × Recall type × Element) ANOVA found significant main effect of concurrent task, F (2, 44) = 72.618, p < 0.001, $\eta_p{}^2$ = 0.767, MSE = 0.588, recall type, F (1, 22) = 9.787, p = 0.005, $\eta_p{}^2$ = 0.308, MSE = 0.206, and element, F (1.14, 25.414) = 32.595, p < 0.001, $\eta_p{}^2$ = 0.597, MSE = 0.074. Element interacted with concurrent task, F (2.38, 52.37) = 5.190, p = 0.001, $\eta_p{}^2$= 0.191, MSE = 0.044, but there was no other two-way or three-way interaction (*ps* > 0.05).

One-tailed dependent *t*-tests with Bonferroni corrections showed significant articulatory suppression effect in movement and colour (*ps* < 0.05), but not in object (p = 0.108) and colourful object (p = 0.064). The significant backward counting effect was significant existed in all elements and colourful objects (*ps* < 0.05). One-tailed independent *t*-test with Bonferroni corrections showed the presence of action advantage in all elements and colourful objects (*ps* < 0.05).

The planned contrast showed that the articulatory suppression disrupted the memory of movement more than that of colour, F (1, 22) = 8.18., p = 0.009, $\eta_p{}^2$ = 0.271, MSE = 0.135, and it also disrupted memory of colour more than that of object, F (1, 22) = 6.168, p = 0.021, $\eta_p{}^2$ = 0.219, MSE = 0.019. Backward counting disrupted recall of colour and movement similarly, F (1, 22) = 3.893, p = 0.081, $\eta_p{}^2$ = 0.061, MSE = 0.146, and had similar effect on the memory of colour and object, F (1, 22) = 0.081, p = 0.779, $\eta_p{}^2$ = 0.004, MSE = 0.013.

Binding

The method of testing the involvement of central executive in binding was same as that in previous experiments. The disruptive effect of backward counting in binding a colour to an object was tested using a 2 × 2 × 2 (Backward counting × Binding × Recall type) ANOVA, the variable binding included colourful object and object, and the variable backward counting included articulatory suppression conditions and backward counting conditions. Results showed significant main effect of backward counting, F (1, 22) = 94.613, $p < 0.001$, $\eta_p{}^2 = 0.811$, MSE = 0.349, binding, F (1, 22) = 30.803, $p < 0.001$, $\eta_p{}^2 = 0.583$, MSE = 0.008, and recall type, F (1, 22) = 9.590, p = 0.005, $\eta_p{}^2 = 0.304$, MSE = 0.247. There was no significant interaction between backward counting and binding, F (1, 22) = 1.766, $p = 0.196$, $\eta_p{}^2 = 0.075$, MSE = 0.004, and no other two-way or three-way interaction (*ps* > 0.05).

The effect of backward counting in binding a movement to corresponding object was tested using a 2 × 2 × 2 (Backward counting × Binding × Recall type) ANOVA, the variable binding included colourful object and action. Results showed significant main effect of backward counting, F (1, 22) = 82.728, $p < 0.001$, $\eta_p{}^2$ = 0.790, MSE = 0.384, binding, F (1, 22) = 240.991, $p < 0.001$, $\eta_p{}^2 = 0.916$, MSE = 0.028, and recall type, F (1, 22) = 10.971, $p < 0.001$, $\eta_p{}^2 = 0.333$, MSE = 0.235. There was no significant interaction between backward counting and binding, F (1, 22) = 1.096, $p = 0.306$, $\eta_p{}^2 = 0.047$, MSE = 0.035, and no other two-way or three-way interaction (*ps* > 0.05).

Serial positions

Each position was coded for the percentage of correct trials, ranged from 0 to 1. A 3 × 2 × 5 (Concurrent task × Recall type × Serial position) ANOVA found significant main effect of concurrent task, F (2, 44) = 63.33, $p < 0.001$, $\eta_p{}^2 = 0.742$, MSE = 7.486, recall type, F (1, 22) = 11.952, $p = 0.002$, $\eta_p{}^2 = 0.352$, MSE = 1.259, and serial position, F (2.13, 46.93) = 127.405, $p < 0.001$, $\eta_p{}^2 = 0.853$, MSE = 5.680. Serial position interacted with concurrent task, F (8, 176) = 10.854, $p < 0.001$, $\eta_p{}^2$= 0.330, MSE = 1.846, and also with recall type, F (2.13, 46.93) = 3.196, $p = 0.047$, $\eta_p{}^2 = 0.127$, MSE = 5.680. There was no other two-way or three-way interaction (*ps* > 0.05).

The post-hoc tests with Bonferroni corrections indicated that the decrease of performance of recall was significant between all adjacent positions (*ps* < 0.05). Comparison of action advantage between adjacent positions indicated that the action advantage was larger in position 3 than in position 2 ($p = 0.037$).

Independent *t*-tests with Bonferroni corrections showed that absence of action advantage in position 1 and 2 (*ps* > 0.100), but evident in all three later positions (*ps* < 0.05). One-way dependent *t*-tests with Bonferroni corrections found significant articulatory suppression effect in positions 3, 4 and 5 (*ps* < 0.05) but not in the first and second position (*ps* > 0.100), and significant backward counting effect in all positions (*ps* < 0.05). The serial position curves as functions of concurrent tasks and type of recall were presented in Figure 5.2.

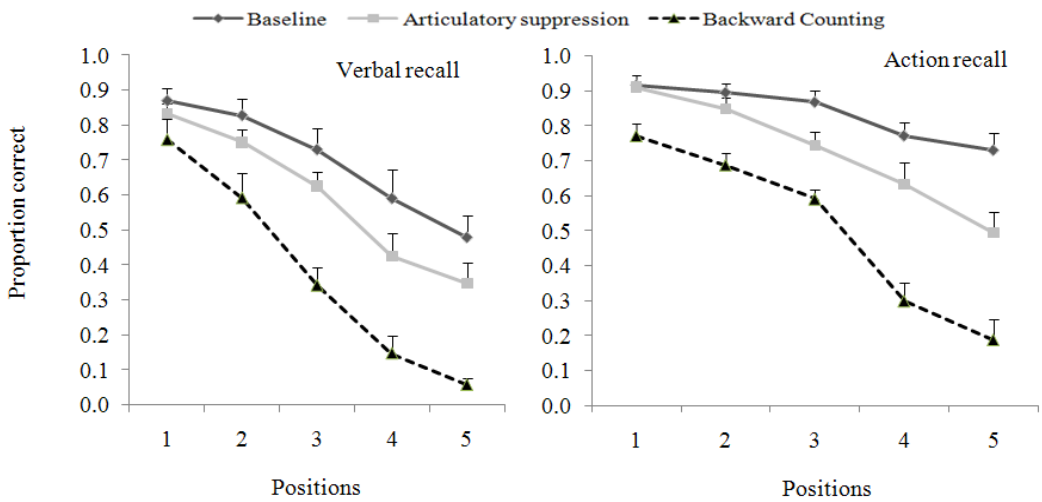

Figure 5.2 The serial position curves (means and standard errors) as functions of concurrent tasks and type of recall in Experiment 6.

Proportion of order errors

The scoring method of proportion of order errors was same as that in previous experiments. The means (and standard deviations) as functions of proportion of order errors as functions of concurrent task and type of recall are shown in Table 5.3.

Table 5.3 Proportion of order errors in Experiment 6

	Baseline	Articulatory suppression	Tapping
Verbal recall	0.05 (0.09)	0.05 (0.08)	0.13 (0.12)
Action recall	0.02 (0.02)	0.05 (0.05)	0.08 (0.10)

A 3 × 2 ANOVA (Concurrent task × Recall type) showed significant main effect of concurrent task, F (2,44) = 4.886, p = 0.012, $\eta_p{}^2$ = 0.182, MSE = 0.006, no significant effect of recall type, F (1, 22) = 2.006, p = 0.171, $\eta_p{}^2$ = 0.084, MSE = 0.003, and no significant interaction between concurrent task and recall type, F (1, 44) = 0.541, p = 0.586, $\eta_p{}^2$ = 0.024, MSE = 0.007. Post-hoc tests with Bonferroni corrections showed that there was more order errors in backward counting conditions than in the articulatory suppression conditions (p = 0.022), but no difference between articulatory suppression and baseline conditions (p = 0.100).

Strategy report

All twenty-four participants reported their use of strategy. The scoring method was the same as that in previous experiments. The count scores and percentage scores as functions of concurrent tasks and type of recall are presented in Table 5.4.

Although the eye-tracking strategy was not provided in the options, observations showed that all participants glanced at the objects frequently when reading the instructions on the screen. The interviews with participants concerning the way in which they read the instructions were informal and are thus not reported here; however, they will be discussed in the Discussion section.

Table 5.4 Self-report strategies in Experiment 6

Verbal Recall	Baseline		Articulatory suppression		Backward counting		Subtotal	
(N = 12)	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Imagining carrying out the actions	12	100%	9	75%	6	50%	28	78%
Remember words visually	2	17%	5	42%	6	50%	13	36%
Verbal rehearsal	8	67%	2	17%	2	17%	12	33%
Grouping	6	50%	4	33%	1	8%	11	31%
Decreasing interference	0	0%	6	50%	3	25%	9	25%
Action Recall	**Baseline**		**Articulatory suppression**		**Backward counting**		**Subtotal**	
(N = 12)	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Imagining carrying out the actions	11	92%	9	75%	10	83%	31	86%
Grouping	7	58%	6	50%	5	42%	18	50%
Verbal rehearsal	7	58%	1	8%	0	0%	8	22%
Decreasing interference	0	0%	6	50%	2	17%	8	22%
Remember words visually	1	8%	1	8%	1	8%	3	8%
Total	**Baseline**		**Articulatory suppression**		**Backward counting**		**Total**	
(N = 24)	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Imagining carrying out the actions	23	96%	18	75%	16	67%	58	81%
Grouping	13	54%	10	42%	6	25%	29	40%
Verbal rehearsal	15	63%	3	13%	2	8%	20	28%
Decreasing interference	0	0%	12	50%	5	21%	17	24%
Remember words visually	3	13%	6	25%	7	29%	16	22%

Discussion

This experiment investigated the role of the phonological loop and central executive in following written instructions using a dual task methodology. All three hypotheses in this experiment were validated. First, there were significant articulatory suppression and backward counting effects, indicating the involvement of the phonological loop and central executive in following written instructions. This result is consistent with results of the experiment investigating following spoken instructions (Experiment 3) and existing literature (Allen, 2009; Gathercole, et al., 2008). Moreover, consistent with the second hypothesis and the findings of the spoken instruction experiment (Experiment 3), the backward counting effect was larger than the articulatory suppression effect, suggesting a greater contribution from the central executive than from the phonological loop. Third, the performance of recall by actions was better than that of oral repetition, therefore extending the phenomenon of action advantage to following written instructions, consistent with existing literature (Koriat, et al., 1990). Finally, there was no significant interaction between working memory and type of recall, suggesting that the contributions of the working memory to the two types of recall were similar. This lack of interaction was the same as that in spoken instructions, again implying that working memory is unlikely to be the source of the action advantage.

Articulatory suppression disrupted memory of action, suggesting the importance of the phonological loop in following written instructions. Moreover, the involvement of the phonological loop is larger compared to its involvement in spoken instructions, which might be due to the additional work of recoding visual forms of words into the phonological representations (Vallar & Papagno, 2002). Another possibility is that when following written instructions, participants had to keep rehearsing the name of an object before located it in the visual display; and if more than one object were maintained before located, more rehearsal was needed. This conjecture is supported by the shape of serial position curves in action recall conditions, in which the trend of a larger articulatory suppression effect in the later than in the earlier positions can be observed (Figure 5.2), implying a growing amount of rehearsal. This pattern shows a sharp contrast with the pattern in spoken instructions (see right panel in Figures 3.2 and 4.2), in which the articulatory suppression effect is relatively small and constant across positions. This again indicates that more rehearsal was needed in order to locate relevant objects in space when instructions were presented visually; spoken instructions, however, allowed the simultaneous tracking of objects as their names were mentioned, thus lessening the burden of rehearsal.

Besides the need of rehearsing object, articulatory suppression effect was greater in maintenance of motor information than maintaining colour and object. This is consistent with the subject report of strategy in Koriat et al.'s study (1990), in which motor information was preferred to be rehearsed comparing to information of colour and object. One possibility is that the to-be-enacted objects can be mapped onto the series of locations in space, whereas the information of movement was more abstract, thus were more likely to be rehearsed and retained in the phonological store.

The contribution from the central executive was substantial and greater than that of the phonological loop. The backward counting effect was evident in movement, colour, and object, as well as in colourful object, reflecting the involvement of central executive in all types of elements. Moreover, the backward counting effect increased the proportion of order errors, indicating the role of the central executive in encoding and maintaining a sequence. These roles were similar to those found in the spoken instructions. Moreover, as with the spoken instructions, the central executive did not help bind movement and the associated colourful object. Nevertheless, the role of the central executive in binding colour to object in following spoken instructions was not replicated in following written instructions; the backward counting effect was similar in object and colourful object. Therefore, it seems that role of the central executive in binding is not as robust as has been thought, corresponding to the inconsistent findings in the literature (Allen, et al., 2006; Baddeley, et al., 2011; Brown & Brockmole, 2010; Karlsen, et al., 2010). In spoken instructions, a greater backward counting effect on colour and object than on movement was not replicated in this experiment; nevertheless, there was a trend of a larger backward counting effect on colour than on movement. These results seem to imply that the involvement of the central executive in remembering visual features of an object in written instructions was smaller than that in spoken instructions.

As in the spoken instructions, a superior performance of recall by actions than by oral repetition was found in written instructions, suggesting that the type of presentation was not relevant to the occurrence of action advantage. As in the experiments using spoken instructions, the action advantage was found to be present in all types of elements as well as in combinations of colours and objects. The action advantage varied with serial positions; specifically, the action advantage was absent in earlier positions (1 and 2) but present in later positions (see also Figure 5.2). In other words, the benefit of enactment relative to oral repetitions did not show until the middle of the action sequence. It thus seems that action planning during encoding or enactment during retrieval, or both, help preserve the actions later in the sequence. It is further conjectured that the action advantage tends to occur when the instructions are lengthy and involve multiple steps, but less likely to show when instructional message are short and simple.

Another finding related to action advantage is that the concurrent tasks disrupted the performance of recall to a similar extent. The same pattern was present for backward counting, suggesting similar contributions from the central executive in building representations for the two types of recall. Again, these results raise doubts of the dual representation hypothesis; instead, the current results imply that a similar representation was formed for the purposes of both action and verbal recall.

Compared to following spoken instructions, participants had more control over the process by which they encoded and followed written instructions. This flexibility was mainly reflected in the ways of reading instructions and strategies employed during encoding, which tended to vary according to the available cognitive resources and types of recall. The majority of participants indicated that they read in sequence, starting from the first action until the last action, and then re-read them starting from the first action. During the re-reading time, they tended to choose individual actions like ‘spin’, ‘touch’ and ‘push’, which were generally considered more difficult and less distinguishable than the concatenated action ‘pick up…and put it into…’. In the articulatory suppression and backward counting conditions, some participants admitted that they were only able to read until the third and fourth action, which helps explain the sharp decline of serial position curves in these conditions.

The largest difference in the use of strategies between the two recall groups is reflected in the ‘remembering words visually’ strategy, with 36 percent of participants in the verbal recall group using this method in contrast to only 8 percent in the action recall group. It seems that participants took advantage of visual codes of words to retain these words and assist oral repetition, which is common in immediate verbal serial recall tasks (Logie, 2003). By contrast, in preparation for actions, remembering the exact words in an instruction was less important. What is more important for the success of action recall is to represent verbal commands as a spatial representation of future actions by mapping movements onto the associated objects in space. The imaginative process thus needs support from the visuospatial sketchpad, which is better when it is not occupied by unhelpful cognitive processes like coding the visual forms of the words. This conjecture was further supported by the evidence of more participants ‘grouping actions’ and ‘imagining doing it’ in the action recall group than in the verbal recall group, suggesting the early attempt of action planning during encoding. It is thus inferred that, although the visuospatial sketchpad had a similar amount of involvement, it might have contributed differently to the building of representations for the two types of recall.

Experiment 7 examined the contribution of the visuospatial sketchpad to following written instructions, and whether it contributes differently to the construction of representations for the verbal and action recall.

Experiment 7

This experiment aimed to examine the role of the visuospatial sketchpad in encoding written instructions. Previous experiments on following spoken instructions found large disruptive effect of eye-closure on memory of instructions but little impairment from the spatial tapping task. The lacking of tapping effect was inferred to be caused by the simplicity of the task, which was repetitive and gradually became automatic, hence requiring little storage of spatial information and thus producing insufficient interference to the spatial coding. Therefore, a more complex spatial taping task was adapted from a classic spatial span task, the Corsi block task, and was used as the concurrent interference task for the visuospatial sketchpad in this experiment.

The Corsi block task is a commonly-used task that taps the visuospatial short-term memory (Berch, et al., 1998; Corsi, 1972; Milner, 1971; Smyth & Scholey, 1992; Vandierendonck, Kemps, Fastame, & Szmalec, 2004). The original apparatus consisted of a set of nine identical blocks irregularly positioned on a wooden board. The experimenter pointed to a series of blocks at a rate of one block per second, and participants then pointed to the same blocks in their order of presentation. It is a span task, with the block sequences increasing until recall is no longer correct. Later computerized versions presented blocks on a two-dimensional touch screen with the sequence being indicated by the changing colours of the blocks in sequence

(Vandierendonck, et al., 2004). Although it is argued that memorizing sequential information tends to draw upon central executive resources (Frisk & Sharp, 2003; Jones, et al., 1995b; Rudkin, et al., 2007), a recent study using dual task methodology showed that the Corsi block task draws upon the central executive only when the sequence to be recalled is longer than three or four items (Vandierendonck, et al., 2004).

To ensure that the adapted Corsi block tapping task was a pure visuospatial task without the involvement of the central executive, participants were required to tap only three blocks in sequence. To avoid it becoming an automatic procedure memory task, the tapping pattern varied from trial to trial. Early investigations have shown that complex tapping configurations place a greater spatial demand on participants than simple configurations (Berch, et al., 1998; Busch, Farrell, Lisdahl-Medina, & Krikorian, 2005; Smirni, Villardita, & Zappalà, 1983); the tapping patterns were therefore designed in such a way to make sure that a large spatial interference was produced. Previous experiments using spoken instructions found that the encoding of instructions in a rich environment relied on the support of the visuospatial sketchpad. Given the same 3D task environment in this experiment, the adapted Corsi block tapping task is expected to impair the recall.

The second aim of this study is to examine whether the visuospatial sketchpad contributes to the rise of action advantage. In the discussion of the previous experiment, difference in strategies indicated different roles played by the visuospatial sketchpad in the two types of recall. For oral repetition, the visuospatial sketchpad supported the maintenance of visual forms of words, whereas for action recall, the visuospatial sketchpad was involved in action planning. Literature has long suggested that the active action planning that helps link the movement and associated object is a key factor for the rise of the action advantage (Allen, 2009; Gathercole, et al., 2008;

Koriat, et al., 1990). Furthermore, if the support of the visuospatial sketchpad were indeed crucial in planning actions, interfering with its working should lead to the decrease or disappearance of the action advantage. This hypothesis was tested in this experiment.

This experiment also provided an opportunity to compare the contributions of the phonological loop and the visuospatial sketchpad to following written instructions. An articulatory suppression task was used to disrupt the phonological loop, and the adapted Corsi block tapping task was used to impair the spatial coding of the visuospatial sketchpad. The two interference concurrent tasks were made comparable in their memory load. The articulatory suppression task required retaining three digits in the memory and repeating them continuously, and the Corsi block tapping task required maintaining three locations in the memory while continuously tapping them. The rate of articulation and tapping was also made the same.

Three hypotheses were tested in this experiment. First, that spatial coding is important in the process of representing instructions in the 3D task environment; therefore it was expected that the Corsi block tapping task that disrupted this coding would also impair the performance of recall. Second, action advantage was expected in this rich task environment, as in previous experiments. Third, as it is inferred that the visuospatial sketchpad might have contributed to the rise of action advantage, the Corsi block tapping task was thus expected to disrupt the performance of action recall more than the performance of verbal recall.

Method

Participants

A total of 36 native English speakers at the University of York were recruited through the electronic booking system in exchange for course credit or an honorarium of £6. None had taken part in the previous six experiments. There were 28 females and 8 males, aged from 18 to 32, with a mean age of 20.31.

Materials

Lists of instructions were the same as those in the previous experiment. A total of 32 three-digit numbers for the practice and two concurrent tasks were created, with numbers for the tapping condition corresponding to 3 of the 9 locations on the Corsi board. Tapping sequences were created to form relatively large triangles in space, with no tapping sequences involving 3 immediately adjacent locations. Thus, a tapping sequence might involve 3-2-8, but not 3-2-4 (see Figure 5.3). Half of the digits sets required a clockwise tapping pattern and the other half required anticlockwise tapping, randomly intermixed (see Appendix 8).

The arrangement of objects and computer screen was equivalent to the previous experiment, except that a Corsi-block board (28 cm × 23cm) was fixed under the table and hidden from the view of participants (see Figure 5.3). The board was taken from the Block recall subtest in the Working Memory Test Battery for children (Pickering & Gathercole, 2001). The numbers on the blocks were placed in the direction facing the participants to allow fast locations of the blocks during the initial tapping.

Design

In a 3×2 mixed design, concurrent task was the within-subject variable, including baseline, articulatory suppression and tapping conditions; and recall type was the between-subject variable, including verbal and action recall. Measurements were same as those in Experiment 6.

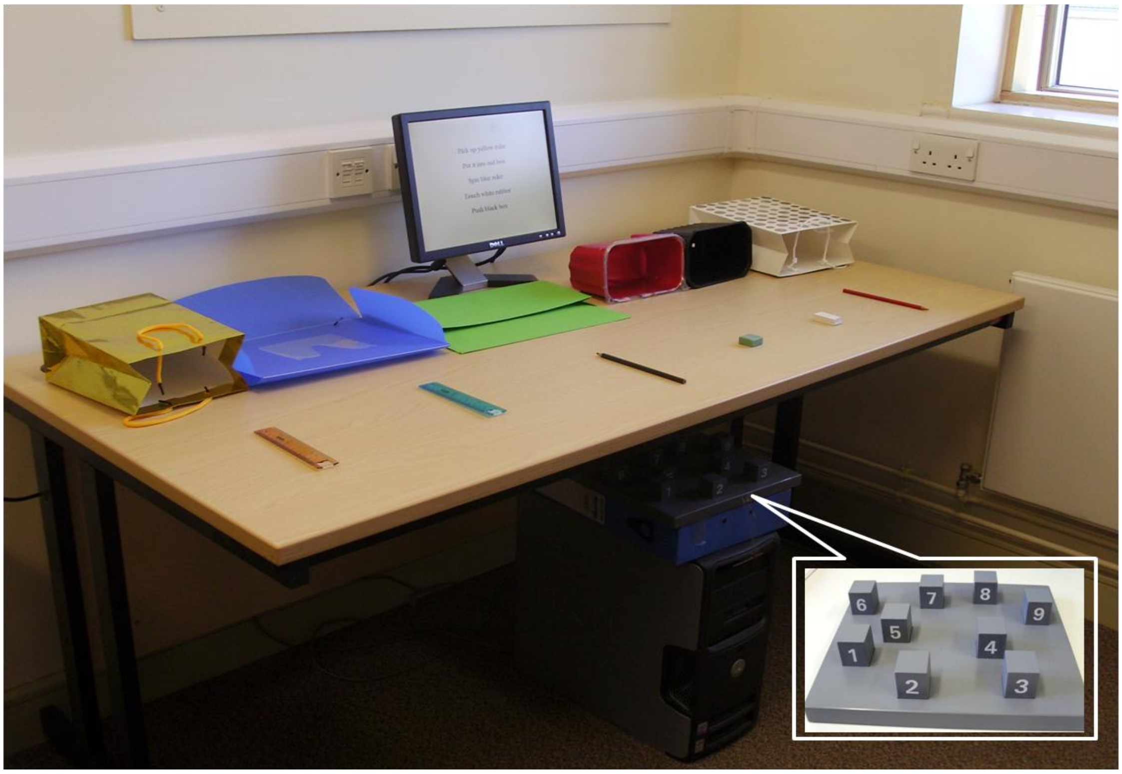

Figure 5.3 Display of 3D instructional task and the Corsi-block tapping board in Experiment 7

Procedure

Each participant completed three conditions. The procedure was equivalent to the previous experiment except that the duration of presenting numbers in the articulatory suppression conditions was changed to four seconds. This was done to make it comparable to the additional time required in the tapping condition to search the tapping blocks and begin the tapping sequence.

In the tapping conditions, upon seeing a three-number digit, the participant first located the corresponding tapping blocks on the Corsi block board and began tapping at the paced rate. Participants were required to tap using a fixed hand configuration (outstretched index finger with the hand shaped as a fist). Participants were allowed to look at the blocks during the first round of tapping, but then had to tap them without looking at them. When instructions were shown on the monitor, participants had to read the instructions and keep tapping at the same time until the beep sound indicating recall.

According to the assigned recall group, participants either verbally reported or physically enacted the instructions. The recording methods were same as those in the previous experiments. At the end of the experiment, the strategy questionnaire was given out.

Results

Actions

All scoring methods were same as previous experiments. Means and the standard deviation of actions and elements are displayed in Table 5.5.

Table 5.5 Means (and standard deviations) of actions in Experiment 7

	Verbal recall	Action recall	Means
Baseline	3.57 (0.81)	4.07 (0.57)	3.82 (0.74)
Articulatory suppression	3.12 (0.58)	3.50 (0.54)	3.31 (0.58)
Tapping	2.59 (0.74)	2.95 (0.76)	2.77 (0.76)
Means	3.10 (0.61)	3.51 (0.48)	3.30 (0.58)

A 3 × 2 (Concurrent task × Recall type) ANOVA revealed significant main effect of concurrent task, F (2, 68) = 41.463, $p < 0.001$, $\eta_p^2 = 0.549$, MSE = 0.239, significant main effect of recall type, action recall was better than verbal recall, F (1, 34) = 5.176, $p = 0.029$, $\eta_p^2 = 0.297$, MSE = 0.268. There was no significant interaction between concurrent task and recall type, F (2, 68) = 0.230, p = 0.795, η_p^2 = 0.007, MSE = 0.239.

Planned contrasts showed significant articulatory suppression effect, F (1, 34) = 22.853, $p < 0.001$, η_p^2 = 0.402, MSE = 0.403, and significant tapping effect, F (1, 34) = 65.113, $p < 0.001$, η_p^2 = 0.657, MSE = 0.607. The articulatory suppression effect did not interact with recall type, F (1, 34) = 0.365, p = 0.550, η_p^2 = 0.011, MSE = 0.403, nor did tapping effect interact with recall type, F (1, 34) = 0.297, p = 0.589, η_p^2 = 0.009, MSE = 0.607.

Scores of the two effects were calculated for each participant. The 2 × 2 (Effect × Recall type) ANOVA found larger tapping effect than articulatory suppression effect, F (1, 34) = 25.153, $p < 0.001$, η_p^2 = 0.425, MSE = 0.211. There was no significant effect of recall type, F (1, 34) = 0.004, p = 0.949, $\eta_p^2 < 0.001$, MSE = 0.211, nor there was any significant interaction between effect and recall type, F (1, 34) = 0.409, p = 0.527, η_p^2 = 0.012, MSE = 0.400.

One-tailed dependent *t*-tests with Bonferroni corrections showed significant articulatory suppression effect in both types of recall (*ps* < 0.05), and also significant tapping effect in both types of recall (*ps* < 0.05). One-tailed independent *t*-tests found no significant action advantage in baseline (p = 0.060), articulatory suppression (p = 0.081) and tapping condition (p = 0.234).

Elements

Each action contained three elements, movement, colour, and object. The scoring method was same as those in previous experiments. As in the previous experiments, the chance levels for elements were different (see Experiment 3 for details).The means and standard deviations of the elements and colourful objects were presented in Table 5.6.

Table 5.6 Means (and standard deviations) of elements in Experiment 7

		Verbal recall	Action recall	Means
Movement	Baseline	3.76 (0.79)	4.25 (0.54)	4.00 (0.71)
	Articulatory suppression	3.38 (0.59)	3.74 (0.54)	3.56 (0.58)
	Tapping	3.07 (0.71)	3.34 (0.79)	3.20 (0.75)
	Means	3.40 (0.61)	3.77 (0.49)	3.59 (0.58)
Colour	Baseline	4.08 (0.83)	4.42 (0.39)	4.25 (0.67)
	Articulatory suppression	3.86 (0.66)	4.00 (0.49)	3.93 (0.56)
	Tapping	3.24 (0.67)	3.40 (0.68)	3.32 (0.67)
	Means	3.73 (0.63)	3.94 (0.41)	3.83 (0.54)
Object	Baseline	4.09 (0.79)	4.48 (0.39)	4.29 (0.65)
	Articulatory suppression	3.83 (0.65)	4.02 (0.44)	3.92 (0.55)
	Tapping	3.31 (0.63)	3.47 (0.66)	3.39 (0.64)
	Means	3.74 (0.60)	3.99 (0.37)	3.87 (0.51)
Colourful object	Baseline	4.03 (0.83)	4.41 (0.39)	4.22 (0.67)
	Articulatory suppression	3.74 (0.64)	3.94 (0.46)	3.84 (0.56)
	Tapping	3.14 (0.67)	3.38 (0.68)	3.26 (0.67)
	Means	3.63 (0.62)	3.91 (0.39)	3.77 (0.53)

A 3 × 2 × 3 ANOVA (Concurrent task × Recall type × Element) showed significant main effect of element, F (1.27, 43.03) = 32.535, $p < 0.001$, $\eta_p{}^2 = 0.489$, MSE = 0.122, significant main effect of concurrent task, F (2, 68) = 39.533, $p < 0.001$, $\eta_p{}^2 = 0.538$, MSE = 0.528, but no significant main effect of recall type, F (1, 34) = 2.685, $p = 0.111$, $\eta_p{}^2 = 0.073$, MSE = 0.262. Element showed significant interaction with concurrent task, F (2.99, 101.52) = 5.670, $p = 0.001$, $\eta_p{}^2 = 0.143$, MSE = 0.036, and there was no other two-way or three-way interaction (*ps* > 0.05).

Dependent one-tailed *t*-test with Bonferroni corrections found significant articulatory suppression effect in all elements and colourful object (*ps* < 0.05) and significant tapping effect in all elements and colourful object (*ps* < 0.05). The planned contrasts showed that the neither of the two effects interact with element (*ps* > 0.05).

Binding

The method of testing the contribution of the visuospatial sketchpad in binding was same as Experiment 4. The effect of tapping in binding a colour to an object was tested using a 2 × 2 × 2 (Tapping × Binding × Recall type) ANOVA, the variable tapping included baseline and tapping condition, and the variable binding included colourful object and object. Results showed significant main effect of tapping, F (1, 34) = 63.857, $p < 0.001$, $\eta_p^2 = 0.653$, MSE = 0.486, binding, F (1, 34) = 24.013, $p < 0.001$, $\eta_p^2 = 0.414$, MSE = 0.014, but no significant effect of recall type, F (1, 34) = 2.680, $p = 0.111$, $\eta_p^2 = 0.073$, MSE = 0.291. The interaction between tapping and binding was significant, with a greater disruptive effect of tapping on colourful object than on object, F (1, 34) = 4.251, $p = 0.047$, $\eta_p^2 = 0.111$, MSE = 0.073. There was no other two-way or three-way interaction (*ps* > 0.05).

The tapping effect in binding a movement to the corresponding colourful object was tested using a 2 × 2 × 2 (Tapping × Binding × Recall type) ANOVA, the variable binding included movement and action. Results showed significant main effect of tapping, F (1, 34) = 69.725, $p < 0.001$, $\eta_p^2 = 0.672$, MSE = 0.521, binding, F (1, 34) = 121.639, $p < 0.001$, $\eta_p^2 = 0.782$, MSE = 0.059, and the effect of recall type was approaching significance, F (1, 34) = 3.788, $p = 0.060$, $\eta_p^2 = 0.100$, MSE = 0.330. The interaction between binding and tapping was not significant, F (1, 34) = 3.003, $p = 0.092$, $\eta_p^2 = 0.081$, MSE = 0.023. There was a significant three-way interaction, F (1, 34) < 0.001, $p = 0.966$, $\eta_p^2 < 0.001$, MSE = 0.023, but no other two-way interaction (ps > 0.05).

Serial positions

Each position was coded for the percentage of correct trials, ranged from 0 to 1. A 3 × 2 × 5 (Concurrent task × Recall type × Serial position) ANOVA found significant main effect of concurrent task, F (2, 68) = 41.715, p < 0.001, $\eta_p{}^2$ = 0.551, MSE = 6.865, recall type, F (1, 34) = 5.121, p = 0.030, $\eta_p{}^2$ = 0.131, MSE = 1.719, and serial position, F (1.93, 65.51) = 121.374, p < 0.001, $\eta_p{}^2$= 0.781, MSE = 9.546. Serial position interacted with concurrent task, F (5.53, 188.04) = 5.206, p < 0.001, $\eta_p{}^2$ = 0.133, MSE = 3.018, but there was no other two-way interaction (*ps* > 0.05). There was no interaction between concurrent task and recall type, F (2, 68) = 0.221, p = 0.802, $\eta_p{}^2$ = 0.006, MSE = 6.865. There was a significant three-way interaction, F (6.69, 188.04) = 2.217, p = 0.048, $\eta_p{}^2$ = 0.061, MSE = 3.018.

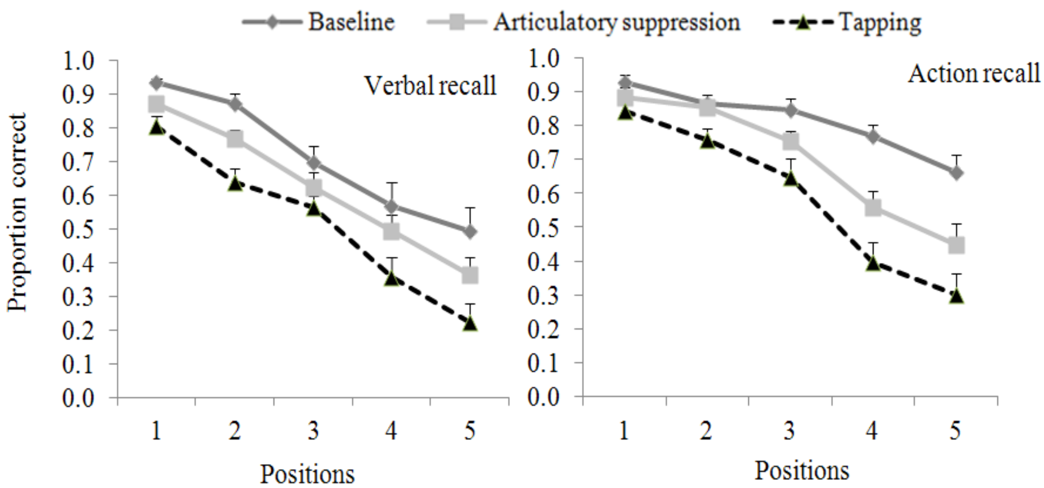

Figure 5.4 The serial position curves (means and standard errors) as functions of concurrent tasks and type of recall in Experiment 7.

Post-hoc test with Bonferroni corrections indicated that the decrement of performance was significant between all adjacent positions (*ps* < 0.001). Independent *t*-tests with Bonferroni corrections showed that absence of action advantage in all positions (*ps* > 0.05) except for position 3 (p = 0.01). Dependent *t*-tests with Bonferroni corrections found articulatory suppression effects were significant in all positions (*ps* < 0.05) except for position 2 (*ps* = 0.058), and tapping effects were significant in all positions (*ps* < 0.05).

Planned contrast showed that articulatory suppression did not interact with any adjacent positions (*ps* > 0.05). The tapping effect interacted with position 1-2, with a greater impairment to position 2 than position1 (p = 0.026), and it also interacted with position 3-4, with greater disruption on position 4 than position 3 (p = 0.008).

Proportion of order errors

The scoring of proportion of order errors was same as that in previous experiment. The means and standard deviations as functions of concurrent tasks and type of recall are shown in Table 5.7.

Table 5.7 Proportion of order errors in Experiment 7

	Baseline	Articulatory suppression	Tapping
Verbal recall	0.04 (0.04)	0.03 (0.04)	0.06 (0.06)
Action recall	0.03 (0.05)	0.03 (0.04)	0.05 (0.07)

A 3 × 2 ANOVA (Concurrent task × Recall type) was conducted. There was no significant main effect of concurrent task, F (2, 68) = 2.260, p = 0.112, η_p^2 = 0.062, MSE = 0.002, no significant main effect of recall type, F (1, 34) = 0.088, p = 0.769, η_p^2 = 0.033, MSE = 0.001, and no interaction between concurrent task and recall type, F (2, 68) = 0.003, p = 0.997, η_p^2 < 0.001, MSE = 0.002.

Strategy report

All twenty-four participants reported their use of strategies. The scoring method was the same as that in Experiment 6. The count scores and percentages of responders as functions of concurrent tasks and type of recall are presented in Table 5.8.

Table 5.8 Self-report strategies in Experiment 7

Verbal recall (N=18)	Baseline		Articulatory suppression		Tapping		Subtotal	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Imagining carrying out the actions	16	89%	14	78%	15	83%	45	83%
Grouping	10	56%	11	61%	9	50%	30	56%
Verbal rehearsal	11	61%	2	11%	6	33%	19	35%
Decreasing interference	0	0%	5	28%	11	61%	16	30%
Remember words visually	1	6%	6	33%	4	22%	11	20%
Action recall (N=18)	**Baseline**		**Articulatory suppression**		**Tapping**		**Subtotal**	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Imagining carrying out the actions	16	89%	15	83%	13	72%	44	81%
Grouping	12	67%	10	56%	7	39%	29	54%
Verbal rehearsal	10	56%	2	11%	8	44%	20	37%
Remember words visually	12	67%	2	11%	1	6%	15	28%
Decreasing interference	0	0%	8	44%	7	39%	15	28%
Use acronyms	1	6%	1	6%	1	6%	3	6%
Total (N=36)	**Baseline**		**Articulatory suppression**		**Tapping**		**Total**	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Imagining carrying out the actions	32	89%	29	81%	28	78%	89	82%
Grouping	22	61%	21	58%	16	44%	59	55%
Verbal rehearsal	21	58%	4	11%	14	39%	39	36%
Remember words visually	12	33%	7	19%	12	33%	31	29%
Decreasing interference	1	3%	14	39%	11	31%	26	24%
Use acronyms	1	3%	1	3%	1	3%	3	3%

Discussion

This experiment focused on the investigation of the contribution of the visuospatial sketchpad to following written instructions and action advantage. There were four main findings in this experiment. First, the Corsi block tapping task had a significantly disruptive effect on performance of recall, indicating the involvement of the visuospatial sketchpad in following written instructions, and thus ratifying the first hypothesis. Second, action advantage was present, establishing the superiority of recall by actions rather than by oral repetition in following written instructions, consistent with Koriat et al. (1990). However, there was no interaction between the tapping effect and recall, disputing the third hypothesis that the Corsi block tapping task would impair action recall more than verbal recall. Finally, the spatial tapping effect was found to be larger than the articulatory suppression effect.

The disruptive effect of the Corsi block tapping task was found to be evident in memory of actions and embedded elements, both individual elements such as movement, colour and object, and also combined entities, like colourful objects. These results indicate that spatial information is important in representing instructions in a 3D task environment. Moreover, the tapping effect was larger in the memory of colourful object than that of objects by themselves, suggesting that the spatial coding might have helped binding colour to an object. These results buttress the argument of using locations as temporary buffers instead of remembering visuomotor details in an action during the course of encoding instructions. Specifically, remembering locations is considered an efficient and economic way of encoding rich visual cues in an environment, as these locations serve as deictic pointers during retrieval (Spivey, et al., 2004). Importantly, these locations might contribute significantly to the glue of multiple features, and disrupting the spatial coding of locations might have forced participants to encode these multiple features separately, probably in a sequentially verbatim way, thus weakening the tight combinations of visual features in locations. If the locations were indeed utilized to form a spatial map of actions, disrupting its function should affect the maintenance of orders, thus leading to the increment of order errors. In this experiment, however, the spatial tapping task did not lead to a significant increment of the proportion of order errors. This might be due to the small proportion and variations of the order errors in this experiment (see Table 5.7).

It was hypothesized the visuospatial sketchpad should benefit the performance of action recall more than the performance of verbal recall. However, the performance of action recall and oral repetition was impaired similarly by spatial tapping indicates the importance of spatial coding in representing instructions for both types of recall. The original hypothesis was inferred from the strategy report in the previous experiment, in which participants in the verbal recall group used ‘remembering the words visually’ more often whereas the participants in the action groups tended to ‘imaging doing the actions’ and ‘grouping the actions’. The latter two strategies were considered related to action planning, which required the visuospatial sketchpad; therefore the interference to the visuospatial sketchpad should diminish the action advantage. However, there was no such difference of the use of strategies between the two recall groups in this experiment; in fact, two groups showed comparable proportions in using these strategies (see Table 5.8). Thus, the visuospatial sketchpad might have involved similarly in representing instructions for future recall, independent of the type of recall. Nevertheless, it should be noted that the two types of recall have different ways of relying on the spatial short-term memory (see Figure 5.4). Specifically, during the action recall conditions, the tapping effect was larger in later serial positions than in earlier ones, suggesting a growing reliance on the spatial storage when coding a sequence. However, the tapping effect remained similar across all serial positions in the verbal recall conditions, indicating a constant reliance on the spatial storage.

The disruptive effect of articulatory suppression on performance of both types of recall was consistent with previous findings, again validating the supporting role of the phonological loop in following written instructions. However, the finding of a larger articulatory suppression effect in memory of movement than in memory of colour and object was not replicated in this experiment, suggesting the argument that motor information was more likely to be rehearsed and stored in the phonological loop was not very robust. Interestingly, similar to the different tapping effect on serial position curves in the verbal and action recall conditions, articulatory suppression also disrupted the encoding processes of the two recall types differently. Specifically, the articulatory suppression effect was constant in all serial positions in the verbal recall conditions, whereas this effect was larger in later positions compared to earlier ones. Therefore, it seems that rehearsal was relied on mainly when remembering actions later in the sequence when the enactment was expected, whereas for oral repetitions, rehearsal was used constantly during the encoding. Together with the finding of the tapping effect, encoding actions for enactment required greater reliance on both storage systems, whereas constant reliance on the two systems could be observed when oral repetitions were expected.

Another finding of this experiment is a relatively larger spatial tapping effect compared to the articulatory suppression effect. Given that the two interference tasks were made similar in their memory load, this result appears to imply greater contributions from the visuospatial sketchpad compared to the contributions from the phonological loop. Nevertheless, it is hard to ascertain whether other aspects of cognitive load in the two tasks were also the same. For example, the spatial tapping task required more attentional control of motor-spatial movement, and may therefore have drawn upon additional executive resources. Therefore, despite a significantly larger spatial tapping effect compared to the articulatory suppression effect, it is risky to conclude that the contribution of the visuospatial sketchpad to remembering written instructions was larger than that of the phonological loop.

General Discussion

All three working memory components were found to be involved in encoding written instructions in these two experiments. Direct comparison of the effect sizes indicated that the greatest contributions came from the central executive, followed by the visuospatial sketchpad, and then the phonological loop. The central executive was found to be related to the maintenance of ordinal information of actions, the visuospatial sketchpad supported the binding of colour to object in an action, and phonological loop has a general supporting role. All three components were involved in coding all types of elements in an action as well as combinations of these.

The phenomenon of action advantage was extended to the situations of following written instructions. Same as that in spoken instructions, the working memory showed similar contributions to both types of recall when instructions were presented visually. It seems that all three working memory components were evoked to build an efficient representation that was useful for later retrieval, no matter which type of recall was required. This representation should not only be efficient in storing all dimensions of elements and facilitating their combinations, but also be effective in retrieving information. A spatial representation of actions is thus an ideal choice. In this representation, the locations are the keys that bind the information from different dimensions: colour, object, and probably also movement. This representation is also a map of routes representing a sequence of actions. In other words, commands of a series of actions were translated to an ordinal path of 'where to do what'. This representation has three benefits. First, it helps offload the cognitive load of remembering visual details of objects to the external world. Second, the map of locations facilitates the binding of the constituents in an action and eases the maintenance of the bound entity. Third, the retrieval process is made easy by simply scanning the mental route of the map and retrieving locations sequentially, from which all detailed information can be extracted. Given these benefits, it is highly likely that such a spatial representation was formed no matter what type of recall was required.

Spoken instructions versus written instructions

It is worth noticing that the findings regarding the involvement of working memory, action advantage, and the lack of interaction between working memory and recall, were similar in following spoken and written instructions.

One difference between the two types of representations is the poorer performance of recall of spoken instructions compared to written instructions. The average of actions per instruction in the baseline conditions across the three experiments using spoken instructions (Experiments 3, 4 and 5) was 3.39, in contrast to 3.83 in those using written instructions (Experiments 6 and 7). There are several reasons underlying the superior performance of recall when instructions were presented visually rather than auditorily. One of the advantages of written instructions is the additional visual codes they offer, which were utilized by the participants to maintain the contents of instructions. Another benefit of the written instructions used in the experiments was their simultaneous presentation of all actions at the same time, which allowed participants to encode instructions at their own pace. This also permitted more flexibility in using strategies during the course of following instructions, such as selectively re-reading difficult action phrases for as long as time allowed. In contrast, the auditory commands decayed rapidly, and once lost, unless rehearsed, the information was lost forever. Given that the instructions in these experiments were lengthy and exceeded the capacity of the phonological loop, the speed of rehearsal might not catch up with the rapid decay of information, leading to the loss of the information. Consequently, in order to keep up with the fleeting oral commands, participants had to draw more on the central executive resource. Indeed, the effect size of the backward counting effect, reflecting the use of the central executive, was larger in the spoken instructions (0.819) than that in the written instructions (0.721).

However, this benefit might be lost or even reversed when instructions are short enough, for which the rehearsal is then sufficient to maintain all the information. Under such circumstances, the extra cost of recoding written instructions into phonological representation, together with the benefit of direct access of auditory commands to the phonological store (Baddeley & Larsen, 2007; Conrad, 1964; Vallar & Papagno, 2002), may lead to an advantage in instructions being presented auditorily than presenting them visually. In addition, delivering instructions auditorily has the benefit of allowing objects to be tracked at the same time as their names are mentioned, thus easing the process of constructing a spatial representation for to-be-enacted objects. Reading written instructions, however, requires a split of attention between reading and looking at a display of objects, leaving fewer attentional resources for encoding of instructions.

In summary, although there were minor differences between the extent to which working memory played a role in following spoken and written instructions, in general, the findings in the spoken instructions have been extended to the written instructions. The next chapter will summarize the findings of seven experiments and discuss the limitations as well as the implications and contributions of this study.

Chapter 6

General Discussion

Overview of thesis

The aim of the thesis was to investigate the contribution of working memory to the process of following instruction sequences. A dual task approach was adopted to identify the roles of the phonological loop, the visuospatial sketchpad and the central executive, the three components in the multicomponent working memory model (Baddeley & Hitch, 1974). Results from seven experiments showed significant involvement of the three working memory components in supporting the encoding of spoken and written instructions, reinforcing the close relationship between working memory and the ability to follow instructions found in literature (Allen, 2009; Engle, et al., 1991; Gathercole, et al., 2008; Kim, et al., 2008).

Moreover, the phenomenon that the performance of recalling instructions by actions was superior to the performance of oral repetition, known as the action advantage, was established in a rich task environment. Although the dual representation hypothesis argues that the action advantage was due to a superior action-based representation for performance versus an inferior verbal-based representation for oral repetition during encoding (Koriat, et al., 1990), all seven experiments in the current study showed a similar involvement of working memory in representing instructions for verbal and action recall. Therefore, the current study implies that the working memory is unlikely to be the source of action advantage.

The following sections review the principal findings, discuss the specific roles of each working memory component as well as the nature of the action advantage, and consider the limitations of the current study and suggestions for future research. Finally, the implications of this programme of research are discussed.

Findings

Chapter 1 provided a broad review of research related to the cognitive process of following instructions. It was noted that following instructions is a complex activity that involves the cooperation of perception, language, memory and action. Specifically, based on the findings in correlation studies (Engle, et al., 1991; Gathercole, et al., 2008; Kim, et al., 2008) as well as in a series of experiments (Allen, 2009), working memory has been identified as an important factor underlying the course of following instructions. A key phenomenon in the process of recalling instructions is the action advantage, that is, the benefit of recalling instructions by actions than by oral repetition. This benefit may be driven by the use of different representations formed for the two types of recall, a verbal-based representation for repetition and an imaginative-based representation for actions (Allen, 2009; Gathercole, et al., 2008; Koriat, et al., 1990). It was thus inferred that constructing different representations should be reflected in the different levels of involvement of the working memory subcomponents. The phonological loop should contribute more to a verbal representation, whereas the visuospatial sketchpad may be more involved in the formation of an action-based representation. These findings and speculations thus formed the two aims of this thesis. The first of these was to investigate the role of working memory in following instructions. The second aim was to explore the action advantage and how working memory contributes to this.

In Chapter 2, a computer-based instructional task was used. This required clicking and dragging using a mouse device to actualize the movement of geometric shapes on a computer screen. After hearing a command like ‘click the flag, drag the star onto the triangle, click the arch, drag the chevron onto the cross’, participants either recalled this by oral repetition, or used the mouse to click and drag to actualize the motion of the shapes on the computer screen. Both articulatory suppression and backward counting tasks were found to impair the performance of recall significantly, suggesting that both the phonological loop and the central executive supported the encoding of instructions. However, contrary to literature (Allen, 2009; Gathercole, et al., 2008), there was no action advantage in the computer-based instructional task, which was inferred to be caused by the poverty of visual and motor cues in the computer-based instructional task.

In order to explore the action advantage experimentally, an instruction task that closely corresponded to the paradigms in which it has previously been found to be robust was developed, and this task formed the basis for the remaining experiments in the thesis. The instructions involved sequences of actions being performed on colourful objects in a three-dimensional space. Participants listened to an instruction such as ‘Push the black pencil, and spin the green eraser, and touch the red pencil, and push the blue ruler, and touch the white eraser’, and recalled the instructions either verbally or by actions. In Experiment 3, both articulatory suppression and backward counting tasks were included to investigate the roles of the phonological loop and the central executive. A small disruptive effect of articulatory suppression and a more substantial effect of backward counting was found, suggesting the involvement of both the phonological loop and central executive respectively. Importantly, action recall was found to be superior to verbal recall, although there was no interaction between the recall mode and concurrent task conditions, suggesting that neither the phonological loop nor the central executive were the source of the action advantage.

The contribution of the visuospatial sketchpad in following instructions was investigated in two experiments described in Chapter 4. In Experiment 4, a repetitive spatial tapping task was shown to cause little impairment to the performance of recall of instructions, which was suspected to be caused by the simplicity of the tapping task and the strategic coding of locations of the objects. Therefore, in Experiment 5, the benefit of spatial coding was removed by requiring participants to close their eyes while listening to the instructions. Eye closure impaired recall, demonstrating the importance of the visuospatial support in an instructional task containing rich visual and spatial cues. Both experiments replicated the action advantage and the lack of interaction between working memory and recall, together with the findings in Experiment 3, implying the similar contributions of working memory to the two types of recall in the 3D instructional task.

Chapter 5 set out to extend the findings in the spoken instructions to the written instructions. Significant effects of articulatory suppression and backward counting effect indicated the involvement of both the phonological loop and the central executive in encoding written instructions. A more complex spatial tapping task was used to disrupt visuospatial coding, and this impaired recall performance significantly. This implied the importance of spatial coding in representing a series of actions upon multiple objects spread over a space. The action advantage was also obtained, and there was no interaction between working memory and the type of recall, again suggesting that working memory contributed similarly to the two types of recall.

Taken together, these findings established the importance of working memory in following spoken and written instructions. The benefit of recalling instructions by actions than oral repetition occurred only in a task environment containing rich visual, spatial and motor cues, and it therefore seems unlikely that action advantage can be attributed to the working memory. These results are summarized in Appendices 9, 10 and 11.

Working memory in encoding instructions

In all seven experiments in this study, the dual task methodology was used to isolate the three working memory components in the multicomponent working memory model (Baddeley & Hitch, 1974). The logic behind this is that tasks using the same cognitive components will compete for resources, and therefore simultaneously processing the two tasks will lead to decrement of performance; by contrast, tasks using different components will not (Baddeley, 1986). The investigation focused on the formation of representation of instructions; therefore all interference tasks disrupted the encoding and maintaining stage of instructions without impeding the recall process.

The phonological loop

The phonological loop comprises a phonological store that contains phonological information, which is susceptible to rapid decay as time passes (Baddeley, 2003b). To prevent decay, a rehearsal mechanism is employed, which also has the function of recoding visual forms of verbal materials into the phonological representations.

The involvement of the phonological loop was investigated using the articulatory suppression task known to disrupt rehearsal (Baddeley, et al., 1975). The reduced performance of recall by articulatory suppression thus indicates the involvement of the phonological loop. Across all six experiments that have investigated the phonological loop, articulatory suppression impaired the performance of recall of actions, suggesting the involvement of the phonological loop in encoding instructions, no matter the instructions were presented visually or auditorily, the task environment was computer-based task or three-dimensional environment containing rich visuomotor cues. With the exception of Experiment 3, a consistent disruptive effect of articulatory suppression on the recall elements like movement, colour, shape and object was found. As the instructions in this study were all verbal in nature, the materials had to be perceived before being stored in other forms; therefore, the phonological loop might be the initial buffer that stores all the verbal information. Therefore, the phonological loop might have played a supporting role in retaining the verbal contents of an instruction before a representation was developed. Although significant, the role of the phonological loop was relatively small in magnitude, consistent with participants' reports that rehearsal strategy was employed relatively infrequently.

Two findings were less consistent across the experiments. The first was that the extent to which the encoding relied on the phonological loop varied with the modality of presentation. Direct comparison of effect sizes indicated that the phonological loop was more involved in encoding written than spoken instructions. This might be due to the additional demand of recoding written words into phonological representation in contrast to the direct entrance of the auditory instructions; this advantage is also known as the modality effect. It has been found that this modality effect is often reflected as a large recency effect in the verbal serial recall, in contrast to no recency effect occurring when material was presented visually (Watkins & Watkins, 1980). This modality effect can also be observed in some action recall conditions in this study; compared with the serial position curves of spoken instructions, the serial position curves in written instructions showed a steeper decline and lacked the recency effect. The modality effect also implies that the different processes were involved in coding auditory and visual materials in spite of having the same verbal content, which supports the modality-view (Baddeley, 2000) whilst opposing a unitary view of working memory (Cowan, 1999).

Another point worth mentioning is people's tendency to retain motor information in the phonological loop. For instance, in Experiment 4, the articulatory suppression effect was larger than the tapping effect on memory of movement, but the reversed pattern was found in memory of colour and object. In Experiment 6, a larger articulatory suppression effect was found in the recall of movement than of colour and object. One possible explanation is that, compared with colour and object, the motor information was more abstract, and was therefore difficult to map directly onto the external world. Therefore, before the motor information could be combined into the representation, it had to be retained by the phonological loop. This is consistent with an observation made in an early study, in which participants indicated that they tended to rehearse the movement in contrast to using the visual imagery to encode the sequence of objects (Koriat, et al., 1990). Nevertheless, it is worth noticing that other experiments did not show the trend of a greater role being played by the phonological loop in coding motor information, and thus the current study did not provide strong evidence for this tendency.

In sum, the role of the phonological loop in encoding instructions was relatively stable across different modalities of presentation as well as across different task paradigms. The phonological loop supports the maintenance of all aspects of verbal materials, both individual and combined elements in an action, and the actions themselves.

The visuospatial sketchpad

The visuospatial sketchpad stores visual, spatial and motor aspects of information (Baddeley, 2007). The visual information is passively stored in a visual cache, whereas a spatially-based rehearsal mechanism called the inner scribe encodes the spatial and motor information (Logie, 1995; 2003).

The role of the visuospatial sketchpad was investigated using spatial tapping tasks and eye closure technique in a rich task environment (Experiment 4, 5 and 7). Except for the simple spatial tapping task in Experiment 4, both the Corsi block tapping task in Experiment 7 and the eye closure requirement in Experiment 5 significantly impaired the performance of recall. The disruptive effect of the simple spatial tapping task in Experiment 4 was weak. This might have been due to the simplicity of the tapping task. First, tapping was restricted to a limited space thus diminishing the spatial demand; second, repetitive tapping became automatic and thus became more akin to a procedure memory task. Therefore, the discussion of the contribution of the visuospatial sketchpad focused mainly on the results of Experiments 5 and 7. Given that the eye closure task blocked the encoding of visual and spatial information, whereas the Corsi block tapping task mainly disrupted the spatial coding, the effects of the two tasks were summarized separately.

First, one should consider the disruption caused by eye closure, which represents the benefit of visual support. The eye closure affected recall of actions as well as of all types of elements, indicating that the visual support had helped with the maintenance of all aspect of information. The effect of eye closure was not only larger in the memory of colour and object than memory of movement, but was also larger in the memory of combined colourful object than of object. These results suggested the importance of visual support in coding visual features and facilitating their bindings. Eye closure also led to an increased proportion of order errors, suggesting that the visual support helped in coding and retaining the sequence of actions. The strategy report found that over 60 percent of participants visually tracked the objects in eye-open conditions. These findings imply that the active tracking of objects in space as their names were being mentioned was crucial for the success of recall of instructions. Importantly, these locations served as deictic pointers in space, which were more efficient and economical than encoding detailed visual features individually (Spivey, et al., 2004). These locations were like pigeonholes containing detailed information, and the only memory needed was of their locations, which probably is a strong cue that eases the retrieval process. This is probably the reason why eye tracking of the objects was the most preferred strategy as long as the visual display was available.

Second, the disruptive effect of the Corsi block spatial tapping task in following written instructions should be considered. The disruption to the spatial coding by the spatial tapping led to a large decrement of the performance of recall of actions and all types of elements, again indicating the importance of spatial coding to the memory of motor, spatial and visual information in an action. These results were consistent with the eye-closure effect in spoken instructions. Moreover, the greater spatial tapping effect on colourful objects than on objects suggests that coding of locations of objects is beneficial to the binding of visual features. Although the spatial tapping tasks did not led to a significant increment of order errors, a trend of increment was observable.

Moreover, a special role of the visuospatial sketchpad in following written instructions is to retain the visual forms of the words. This was reflected in the strategy report, in which 29 percent of participants indicated ‘remembering the words visually’. Although this experiment did not provide evidence for the usefulness of this strategy, the benefit of remembering visual forms of verbal material in verbal serial recall has been documented in literature (Logie, 2003).

In summary, the contribution of the visuospatial sketchpad to encoding series of actions upon multiple colourful objects was substantial. It supported encoding of visual, spatial and motor cues in an action, and also helped to bind these elements.

Moreover, it contributed to the maintenance of the sequence of these actions, probably via the eye tracking of the locations of to-be-enacted objects and referring to these locations during retrieval.

The central executive

The central executive is an attentional system that regulates the phonological loop and visuospatial sketchpad, and is assumed to have a range of executive functions (Baddeley, 2007). Of particular relevance to following instructions are the switching of strategies, dividing attention, planning, sequencing and monitoring actions.

The contribution of the central executive to following instructions was investigated in three experiments (Experiments 1, 3 and 6), covering spoken and written instructions as well as the computer-based and three-dimensional task environments. A backward counting task was selected to disrupt the central executive. The task required participants to count a three-digit number backwards in decrements of three or two (a decrement of three in the computer-based task and a decrement of two in the 3D task). As this task overlapped with the articulatory suppression task in the demand of spoken production of number sequences, the contribution of phonological loop was partialled out. Therefore, the difference in performance between the articulatory suppression and backward counting conditions represented the contribution of the central executive.

Several roles of the central executive were evident and consistent across the three experiments. First, backward counting disrupted the performance of recall of actions significantly, suggesting that the central executive plays a substantial role in following instructions. Moreover, the backward counting effect was larger than the articulatory suppression effect, suggesting that the contribution of the central executive in following instructions is greater than that of the phonological loop. It also suggests that following instructions is a complex task that requires high cognitive functions, and is more than the simple maintenance of instructions.

Second, backward counting impaired the memory of all types of elements, with a greater disruption to memory of colour and object than to memory of movement. After excluding the possible difference of difficulty in encoding these elements, it was inferred that more central executive resources are devoted to memorizing visual aspects of information relative to motor information. It has been speculated that the contribution of the central executive in maintaining visual information is via conscious control of eye movements in remembering locations of objects (Postle, et al., 2006). This speculation was supported by the increased proportion of order errors when the function of the central executive was interfered with by the backward counting task.

The recently proposed function of central executive was to bind information from different modalities and from long-term memory (Baddeley, 2000). In the computer-based task, an action contained two elements, movement and shape, whereas in the 3D instructional task, an action contained three elements, movement, colour and object. One consistent finding across the three experiments was the absence of the central executive in binding movement to object. Worth noticing is that the visuospatial sketchpad also did not have such as role in binding movement and object. These results indicate that the process of associating movement with a corresponding object appears to be a relatively automatic process and perhaps runs outside working memory. The binding of colour to object was investigated only in the 3D instructional task. The involvement of the central executive was evident in Experiment 3 when instructions were presented auditorily, but was absent in Experiment 6 when instructions were presented visually. The inconsistent role of the central executive in binding visual features corresponds to the mixed findings in literature (Allen, et al., 2006; Brown & Brockmole, 2010); thus more evidence is needed before a firm conclusion regarding the role of central executive in binding visual features can be drawn.

In summary, the central executive played an important role in following instructions across both spoken and written instructions. Specifically, the central executive encoded all types of elements and contributed to the sequencing of actions.

The action advantage

One phenomenon observed in previous research is a superior performance of recall by actions than by oral repetition, namely, the action advantage, which exists in both children and adults and in following spoken and written instructions (Allen, 2009; Gathercole, et al., 2008; Koriat, et al., 1990).

Acquiring action advantage

In the initial investigation using the computer-based task, the instructions involved clicking and dragging geometric shapes using the mouse device. The performances of recall by actions and by oral repetition were found to be similar. It was inferred that the lack of action advantage might be due to the poverty of the task environment. Therefore, a 3D instructional task embedded with rich cues was developed, in which participants were required to remember instructions of a series of different actions upon colourful objects displayed in a large space. In this rich environment, a robust action advantage was obtained and replicated in situations when participants were required to follow both spoken instructions (Experiments 3, 4, 5) and written instructions (Experiments 6 and 7). Given the many differences between the computer-based and 3D instructional tasks, it is hard to ascertain any specific factor that gave rise to this action advantage. Nevertheless, it is suspected that the richness of visual and motor cues in the three-dimensional environment might have contributed to this benefit.

Moreover, the benefits of preparing for action recall manifested the improvement of memory of all types of elements in an action. That is, the performance of recall of movement, colour and object was superior in the action recall conditions compared to the verbal recall conditions. This is consistent with the finding that the presence of action advantage in the number of correct features in the classroom instructional task in children (Gathercole, et al., 2008). Thus it seems that the benefit of planning for actions improved the memory of all dimensions in an action. In addition, the finding of the existence of action advantage in combined entities (colourful objects) implies that planning for actions also facilitated combinations of elements in an action. This is consistent with the notion that preparing to perform actions induced extraction of relational aspects, therefore making unitary codes more interactive with each other (Marschark, Richman, Yuille, & Hunt, 1987).

Working memory in action advantage

Previous research has suggested that the action advantage rises in the encoding stage, with a superior imaginal-coded representation for actions compared to a verbal-code representation for oral repetition (Koriat, et al., 1990). Specifically, the action representation integrates multidimensional information and allows access to various perspectives of information simultaneously, whereas the verbal representation is constrained by sequential processing. Support for this dual representation hypothesis was provided by several later studies (Allen, 2009; Gathercole, et al., 2008; Wojcik, Allen, Brown, & Souchay, 2011).

It is worth noticing that, in the Gathercole et al.'s (2008) study, short-term memory and working memory were found to correlate with performance of actions, but not with performance of oral repetition. However, in Allen's experiments, the articulatory suppression and backward counting effect existed in both types of recall, suggesting that working memory is involved in the representations for actions as well as for oral repetition (Allen, 2009). Similarly, across the seven experiments in the current study, working memory was involved in representations for both verbal and action recall. The discrepancy of the involvement of working memory in verbal recall might be due to the different tasks used in these studies. In Gathercole's study, the instructional task was a span task which started with a short sequence of actions, which was within the capacity of the phonological loop and therefore encouraged the verbal coding for oral repetition in the first place. In contrast, the instructional tasks in the current study and Allen's research used a fixed length of instructions that were likely to go beyond the capacity of the phonological loop; as a consequence, all working memory components would have to do all their possible to be able to maintain the instructions in memory, no matter what type of recall was required.

Under circumstances in which lengthy and complex instructions have to be remembered and recalled, an integrated multimodal representation might be the best solution, even for oral repetition. Therefore, it appears that the type of representation form for recall depends on whether the instructional message can be held within one's working memory capacity. If the instruction is relatively short and easy, verbal codes will be used, as these are sufficient for immediate repetition and no translation cost is involved. In contrast, when an instructional message is lengthy and complex, it is better mapped into a multimodal representation that allows simultaneous access of various dimensions of information. Moreover, the external environment can be combined into the representation; for instance, locations can be used as temporary caches, and these deictic points can then be looked back to in order to retrieve detailed visual information (Spivey, et al., 2004).

In summary, it is perhaps better to give short instructions within people's working memory capacity, so people can easily hold them in the phonological loop while carrying out the actions. If the instructions are unavoidably lengthy, people should be taught to link the operations with to-be-performed objects and utilize the surrounding environment in order to offload the burden of maintaining all the detailed information of a series of actions in the working memory.

The verbal output disadvantage

Until now, it has been assumed that the action advantage arises mainly from the encoding stage (Koriat, et al., 1990; Saltz & Dixon, 1982). Nevertheless, Koriat also noticed there were more repetition errors in verbal recall than in action recall, implying a weak output monitoring in the course of oral repetition. In other words, action advantage can be considered to be a verbal output disadvantage. This is because visual features are bound in an object in the multimodal representation (Luck & Vogel, 1997), and oral repetition demands separation of the visual features into sequential outputs; this de-binding process requires attention and hence impairs the recall of actions in the later sequence (Singer & Gray, 1995). In contrast, the features in intended objects are always bound together during action execution.

This de-binding cost explanation is supported by the differences in the serial position curves between the two recall groups. The performance of oral repetition dropped significantly from the first position down to the last position, whereas the serial positions in the action recall showed little decline in the first positions and levelled off until the end. The sharp decline in verbal recall might have been due to the attention devoted to the decoding of bound entities early in the positions, leading to the loss of information later in the sequence. The de-binding cost account also offers an explanation for the lack of action advantage in the computer-based task, in which objects contained only single visual feature; without the de-binding process, oral repetitions became as easy as executing actions. In the 3D task, by contrast, each action contained an additional colour dimension, which required de-binding in the oral repetition, thus leading to the disadvantage of verbal recall.

The contrasts in the serial position curves between verbal and action recall can also be interpreted in terms of the greater output interference in oral repetition versus enactment. That is, a person's own verbal output tends to interfere with his or her representations of items yet-to-be recalled, which is considered to be a major contributor to the rise of primacy effect in verbal serial memory (Cowan, et al., 2002; Oberauer, 2003). In contrast, action output has no such interference; rather, the action output manifested itself as the completion status of objects, serving as reminders of the progress during the course of execution.

Excluding other factors

The current research also provides evidence for ruling out factors that have not contributed to the rise of action advantage. For examples, actions may have more direct and visible goals, and are thus likely to evoke more active processing than oral repetition. In the current study, the primacy effect of the serial position curves, which reflects the active encoding of actions, were evident in both types of recall. Moreover, more active coding should manifest itself in a greater involvement of the central executive; however, the contributions of the central executive were similar in both verbal and action recall. These results thus suggest equal effort and motivation in encoding instructions for different types of recall, thus excluding the possibility that the action advantage was caused by a greater motivation when preparing for actions than for oral repetition.

The possibility that superior strategies were adopted for action performance than for oral repetition was also examined. The strategy reports showed no consistent pattern of difference in strategies between verbal recall and action recall. Moreover, the action advantage cannot be attributed to the difficulty of sequencing actions during oral repetition, as there was no significant difference in the proportion of order errors between the two types of recall across the five experiments that investigated order errors.

In summary, the benefit of recalling instructions by action than oral repetition was established for both spoken and written instructions. Working memory was shown to make similar contributions to the two types of recall. It is speculated that, in order to cope with the working memory demand of coding lengthy instructions, a superior multimodal representation was formed for both types of recall, and the verbal output disadvantage might be one of the reasons that led to the poor performance of oral repetition compared to action performance.

Sequential representations of actions

Although not the primary aim of the study, the serial position curves are informative in revealing the way we remember a series of actions and retrieve them from memory. One heated debate in the area of serial memory is whether it is modality-independent (Depoorter & Vandierendonck, 2009; Jones, et al., 1995b) or not (Smyth, et al., 1988), which is testing whether the primary serial memory task is affected by a secondary serial memory task which processes materials from a different modality. In this study, all concurrent tasks – articulatory suppressions, spatial tapping, and backward counting – involved an ordinal component, and their disruptions to the verbal serial memory of actions thus seem to reflect an independent order system. However, it should be noted that the different concurrent tasks interact with serial positions differently, suggesting that they also have own separate ordinal systems. It is thus speculated that, although the sequence of action commands was presented verbally (be it spoken or written), it may be represented as a sequence of small multimodal episodes. This is consistent with the Burgess and Hitch model (1999), in which the order was coded by associating items with contextual representation that containing multiple layers. Exactly how different layers of information are combined corresponds to the big questions of binding, which is assumed to be the role of the episodic buffer (Baddeley, 2000). For each serial position, occurred information of parallel sequences from different domains can be simultaneously bound to the same contextual signal (Hurlstone, 2010). The merge of sequences, however, is actualized by selective attention directed by the goals of learning (Keele, et al., 2003). In this study, the important contextual signal is likely to be the location, which glues all dimensions of information together. The sequence of the locations in space thus was used to represent the multiple sequences of information, such as movements and object features.

Another important and consistent finding across the seven experiments is the contrasting shapes of serial position curves in verbal and action recall, implying that different cognitive processes were underlying the way in which people represented a sequence of actions. As was discussed in the section on verbal output disadvantage, the spoken output of actions earlier in the list tended to impede the memory of later actions; whereas there was no such proactive interference when enactment was required. It is further conjectured that the absence of this interference during enactment can be potentially more beneficial when instructions are lengthy, and thus vulnerable to proactive interference.

Finally, the serial position data also provides support for the existing effects in the verbal serial memory, including the primacy effect as well as the modality effect. To be specific, the larger recency effect in an aurally- than in a visually-presented sequence (Watkins & Watkins, 1980), was also extended to situations where a series of action commands were remembered and when enactment was required.

How do we follow instructions?

Taking together the findings from objective measurements and strategy reports, as well as the literature, the cognitive process of following instructions can be inferred. In a situation where a verbal instruction involves multiple actions upon objects dispersed in a large space, the task is more than a simple retention of verbal materials. Rather, following instructions is a complex task that requires working memory.

The encoding stage should be considered first. Verbal instructions are perceived and retained by the phonological loop, which allows direct access of auditory command, whereas written instructions need recoding. These phonological codes are maintained before a multimodal representation can be developed. When instructions are lengthy and beyond the capacity of the phonological loop, a multimodal representation is formed that allows multi-dimensional information to be combined and stored efficiently.

Maintenance of the visual information in an action relies mainly on the visuospatial sketchpad, which facilitates the binding of the visual features in an object, sometimes with the help of the central executive. Whenever a visual display of an object is available, people actively eye-track these to-be-enacted objects in sequence as their names are mentioned. These locations are used as deictic pointers and caches for objects and their visual features, which is more cognitively economical than remembering colour and object separately (Spivey, et al., 2004). This eye tracking behaviour also contributes to memorizing the orders of actions, which requires the conscious control of the central executive. As can be seen, the process of building representation is rather complicated and requires the cooperation of various cognitive functions. In particular, the central executive plays a substantial role, probably helping coordinate the two storage components as well as allocating and shifting attention between internal goals and the external world where actions are about to take place.

During retrieval, in the action recall, the multimodal representation can be directly mapped onto the external world by execution. Oral repetition, however, requires costly translation and de-binding of the multimodal representation into a sequential verbal output, which creates interference thus impairing the performance of repetition.

When instructions are short the phonological loop is sufficient to hold the entire commands; a verbal-code based representation is therefore likely to be formed for oral repetition, which is intrinsically inferior to an imaginal action-based representation for actions. The use of the same type of codes to represent and retrieve information helps to prevent the translation cost if different codes were used. In such cases, it is the difference in the quality of representations that led to action advantage (Gathercole, et al., 2008; Koriat, et al., 1990).

It therefore appears that how people represent instructions is related to both the cognitive demand of the task and the goal of the task (oral repetition or performance). This is consistent with the idea that the extent to which working memory is employed depends on the cognitive load of the task (Logie, 2011). People are also flexible in coping with difficult situations. For example, in the eye-closure experiment, when a visual display was not available, participants immediately shifted to rely on other strategies, such as rehearsal and imaging themselves doing the actions. Whenever they were allowed to see the visual display, they tracked the objects in space as the names of these were mentioned. It is as if they knew that eye tracking helped offload the burden of remembering the visual details of an object (Spivey, et al., 2004). Moreover, in all experiments, most participants reported using several strategies rather than relying on a single strategy, suggesting their conscious effort in coding information as a multidimensional representation for later recall. This tendency to utilize several strategies to cope with a complex task is compatible with the view that the use of several mechanisms is usually less taxing than relying on only one mechanism (Cowan, 1988, 1999).

Nevertheless, some aspects of following instructions remain unclear. One is how movement is linked to a corresponding object. One possibility is that the movement is bound to an object via imaging oneself doing the action in an early stage of encoding. Another is that the sequence of movement is maintained in the phonological loop, and is only retrieved and combined with the object at the moment of execution. Unfortunately, the experiments in this study did not help distinguish the two hypotheses. Nevertheless, it is certain that the working memory did not help in binding movement and object. Another unknown cognitive process relating to following instructions is the retrieval stage. Specifically, how working memory supports retrieval of instructions has not been investigated; nevertheless, this may provide important insights into the rise of the action advantage.

Future research

Future research can focus on both theoretical exploration and extension to applied research. I will first consider potential research on understanding the process of following instructions. Current research has established the significant contribution of the working memory in remembering instructions, but the support it provides for the retrieval of instructions remains unknown. Importantly, the findings implied a verbal output disadvantage, which might be related to the cost of translating and de-biding the multimodal representation to the sequential verbal codes. These costs may be reflected in the involvement of working memory; investigating the contribution of working memory in retrieving instructions may therefore help provide evidence for the verbal disadvantage hypothesis.

In this study, frequent eye tracking of objects during encoding was observed, suggesting the importance and benefit of using locations as deictic pointers and caches for to-be-enacted objects. This conjecture can be tested by comparing the patterns of eye-movement during the encoding and retrieval stages (Spivey, et al., 2004). The eye-movement data can also provide insights into the contribution of eye-movement to reading written instructions, in encoding the actions, and performing the actions (Land & Hayhoe, 2001). The difference of eye movement patterns may help explain the rise of the action advantage. For example, eye movement patterns may be more predictable than reactive during the course of the executing of actions.

Although the strategies employed by participants were reported in the study, how they contributed to the performance of recall remains unclear. Knowing the relationship between the different strategies and the ability to follow instructions can be helpful in choosing useful tactics. Moreover, the relations between strategy and the disruptive effect of different tasks can also be informative. For instance, in Experiment 6, the percentage of participants imagining carrying out the actions during encoding was decreased when the tapping was required at the same time, suggesting that the mental simulation of actions was using a similar cognitive function of the tapping. The limited number of participants and scarce reports of strategies in this study prevented an in-depth investigation of this question; nevertheless, future study should consider this important question.

In the same vein, the patterns of errors made under different disruptive tasks should also be examined in a later study. This study only investigated order errors, but not item errors such as omissions, repetitions, and intrusions. The contrasts of error patterns as the result of different disruptive tasks may provide insights into the roles played by the different working memory components. In addition, analyzing the pattern of errors can help unveil the common mistakes in the course of following instructions, thus avoiding these pitfalls, and consequently improving the performance of recall.

Another interesting research topic would be to see whether mental imagination during encoding has any benefits for the performance of recall. A recent study has shown the benefit of subject-performed task in following instructions, which was attributed to the mental practice that reinforced the multimodal representation during encoding, or the benefit of motor coding, or perhaps both (Wojcik, et al., 2011). Therefore, it would be theoretically and practically interesting to disentangle the two contributors by comparing the conditions of mental practice and actual performance. If mental practice indeed has benefits for the recall of instructions, this benefit can be applied to various learning scenarios.

The present research investigated situations in which instructions were spoken and written; however, the process of following the demonstration of actions has not yet been investigated. Previous studies on imitation suggest the existence of a direct mapping of observed actions and imitative actions (see section of direct mapping in

Chapter 1). It is thus inferred that this direct mapping may provide some advantage of recalling by actions than orally describing the actions, which should conceivably be reflected in a decreased involvement of working memory in action recall.

In this study, the objective measurements focused on the accuracy of performance. Future research can also examine the time course of the process of following instructions, such as the preparation time before recall and the duration of recall. These indexes may help depict the time course of following instructions. The time courses in the two types of recall may also elucidate the rise of the action advantage. Furthermore, the instructions in this research contained arbitrary steps of actions rather than a series of actions leading towards a meaningful goal. This was done to mimic the situations of learning new sequence of actions, and focused on the memory process rather than language comprehension. Future studies could investigate instructions containing linked actions, and examine the contribution of schemas of the long-term memory in remembering instructions.

In the area of applied research, the present study could also inspire future research. For example, it is useful to know the development of children's abilities in following teachers' commands. Both teachers and parents can then give children appropriate orders within their working memory capacities. In the case of elderly people, it is worth knowing their difficulty in the course of remembering instructions and during the execution of these. This allows helpful techniques to be developed for following instructions, which may improve their memory of instructions and benefit their daily life. For example, remembering locations of intended objects and imaging themselves doing the actions are both helpful tips. These applied studies can also provide advice for the design and optimization of instructions, and eventually benefit learning.

Conclusions and contributions

This study aimed to fill the gaps in the previous studies that investigated the role of following instructions (Allen, 2009; Gathercole, et al., 2008). Specifically, the two issues remained unclear is the cognitive process of following instructions and the underlying mechanism of the action advantage. This study set out to answer the two questions from the perspective of working memory.

The results showed that working memory was highly involved in the process of following instructions. Central executive had the greatest contribution and was related to direct eye movement to help retain the sequence of actions upon to-be-enacted objects. The phonological loop played a general supporting role in retaining the verbal materials in the phonological store and preventing the decay of information via constant rehearsal. The visuospatial sketchpad helped bind the visual features within an object, probably by means of maintaining a spatial representation of actions A superior recall of actions to oral repetition was established in a rich task environment using both spoken and written instructions, corroborating the phenomenon of the action advantage suggested in literature (Allen, 2009; Gathercole, et al., 2008; Koriat, et al., 1990). However, there was no interaction between working memory and recall, suggesting that the source of this action advantage was unlikely to be in the working memory.

These findings not only establish the involvement of working memory in following instructions, but also provide insights into the roles played by the three working memory components. Moreover, it is the first study which has tried to explain the action advantages in terms of the functioning of working memory during encoding. It is also the first study to have investigated situations in which instructions were presented as written words. Importantly, various aspects of objective measurements, such as elements, binding, serial positions, as well as strategy reports, were included in this study. In particular, the data of serial position provides important insights into the way in which people represent a series of actions sequentially. Taken together, this study helps to depict a comprehensive cognitive process of following instructions, as well as raising intriguing questions for theoretical and applied research in the future.

Finally, in educational situations such as teaching and learning, the current study also provides useful suggestions. For example, teachers should bear in mind that following lengthy instructions can be cognitive demanding; as this study has shown, even for an adult with an undergraduate education, a command including series of five new actions places a large burden on the working memory. Therefore, it is beneficial to divide lengthy instructions into short ones. An awareness of the heavy demand of instructions is especially important to those who have a relatively lower working memory capacity than typical adults, such as children and elderly people, and also clinic populations. The robust action advantage established in this study also hints that we should go on first-hand experience when following instructions.

Appendices

Appendix 1: Lists of instructions in Experiment 1 and 2

1. click the flag drag the star onto the triangle click the arch drag the chevron onto the cross
2. drag the arch onto the chevron click the circle click the cross drag the diamond onto the triangle
3. drag the star onto the flag drag the diamond onto the cross click the circle click the chevron
4. click the triangle click the arch drag the circle onto the flag click the cross click the diamond
5. drag the diamond onto the star click the chevron drag the arch onto the flag click the triangle
6. click the chevron click the cross click the flag drag the triangle onto the arch click the star
7. click the star drag the diamond onto the circle drag the chevron onto the cross click the flag
8. drag the diamond onto the triangle click the flag click the circle drag the star onto the chevron
9. click the arch drag the triangle onto the cross drag the star onto the circle click the diamond
10. drag the flag onto the diamond drag the chevron onto the triangle drag the star onto the cross
11. click the chevron click the circle drag the arch onto the cross drag the triangle onto the star
12. drag the cross onto the diamond click the flag drag the chevron onto the circle click the arch
13. click the star click the triangle click the flag drag the cross onto the chevron click the circle
14. click the circle drag the arch onto the diamond drag the star onto the triangle click the chevron

Appendix 2: Strategy questionnaire in Experiment 1

Following instructions

Gender Male ☐ Female ☐
Age ____
Department ____________________

Condition 1

Please circle the level of difficulty in this experiment

1= very easy 2= slightly easy 3 = moderate 4 = slightly difficult 5 = very difficult

Remember the instructions	1	2	3	4	5
Repeat the instructions in orders	1	2	3	4	5

If you are using any strategy, please state

____

Condition 2

Please circle the level of difficulty in this experiment

1= very easy 2= slightly easy 3 = moderate 4 = slightly difficult 5 = very difficult

Remember the instructions	1	2	3	4	5
Perform out the instructions in orders	1	2	3	4	5

If you are using any strategy, please state

____

Condition 3

Please circle the level of difficulty in this experiment

1= very easy 2= slightly easy 3 = moderate 4 = slightly difficult 5 = very difficult

Remember the instructions	1	2	3	4	5
Repeat the instructions in orders	1	2	3	4	5
Repeat the number when listening to the instructions	1	2	3	4	5

If you are using any strategy, please state

____

Condition 4

Please circle the level of difficulty in this experiment

1= very easy   2= slightly easy   3 = moderate   4 = slightly difficult   5 = very difficult

Remember the instructions	1	2	3	4	5
Perform the instructions in orders	1	2	3	4	5
Repeat the number when listening to the instructions	1	2	3	4	5

If you are using any strategy, please state

____

Condition 5

Please circle the level of difficulty in this experiment

1= very easy   2= slightly easy   3 = moderate   4 = slightly difficult   5 = very difficult

Remember the instructions	1	2	3	4	5
Repeat the instructions in orders	1	2	3	4	5
Backward count the number when listening to the instructions	1	2	3	4	5

If you are using any strategy, please state

____

Condition 6

Please circle the level of difficulty in this experiment

1= very easy   2= slightly easy   3 = moderate   4 = slightly difficult   5 = very difficult

Remember the instructions	1	2	3	4	5
Perform the instructions in orders	1	2	3	4	5
Backward count the number when listening to the instructions	1	2	3	4	5

If you are using any strategy, please state

____

Note. Condition1 refers to the Baseline_verbal recall; Condition 2, Baseline_action recall; Condition 3, Articulatory suppression_verbal recall; Condition 4, Articulatory suppression_action recall; Condition 5, Backward counting_verbal recall; Condition 6, Backward counting_action recall.

Appendix 3: Lists of instructions in Experiment 3

<u>List1</u>

1 Push the black pencil and spin the green eraser and touch the red pencil and push the blue ruler and touch the white eraser

2 Touch the red pencil and push the yellow ruler and pick up the green eraser then put it into the black box and spin the blue ruler

3 Pick up the green eraser then put it into the white bag and push the yellow ruler and pick up the red pencil then put it into the blue folder

4 Spin the blue ruler and touch the green eraser and pick up the black pencil then put it into the yellow bag and touch the white eraser

5 Pick up the white eraser then put it into the green folder and spin the yellow ruler and push the white eraser and touch the red pencil

6 Touch the yellow ruler and spin the red pencil and push the blue ruler and pick up the black pencil then put it into the blue folder

7 Push the black pencil and pick up the yellow ruler then put it into the white bag and touch the green eraser and push the red pencil

8 Pick up the white eraser then put it into the black box and spin the blue ruler and pick up the black pencil and put it into the green folder

9 Touch the green eraser and spin the yellow ruler and pick up the white eraser then put it into the red box and push the black pencil

10 Touch the red pencil and push the blue ruler and spin the white eraser and touch the yellow ruler and push the black pencil

11 Pick up the yellow ruler then put it into the white bag and touch the blue ruler and pick up the white eraser then put it into the red box

12 Push the green eraser and pick up the black pencil then put it into the yellow bag and touch the red pencil and spin the blue ruler

13 Spin the blue ruler and push the green eraser and touch the red pencil and pick up the white eraser then put it into the yellow bag

14 Spin the red pencil and pick up the blue ruler then put it into the green folder and touch the black pencil and push the green eraser

<u>List 2</u>

1 Push the green eraser and touch the black pencil and pick up the blue ruler then put it into the green folder and spin the red pencil

2 Pick up the white eraser then put it into the red box and touch the black pencil and pick up the yellow ruler then put it into the white bag

3 Spin the yellow ruler and pick up the white eraser then put it into the red box and push the black pencil and touch the green eraser

4 Pick up the black pencil and put it into the green folder and spin the blue ruler and pick up the white eraser then put it into the black box

5 Touch the green eraser and pick up the black pencil then put it into the yellow bag and touch the blue ruler and spin the white eraser

6 Push the red pencil and touch the green eraser and pick up the yellow ruler then put it into the green folder and push the black pencil

7 Pick up the black pencil then put it into the blue folder and spin the red pencil and push the blue ruler and touch the white eraser

8 Push the white eraser and touch the red pencil and spin the yellow ruler then pick up the white eraser then put it into the green folder

9 Push the black pencil and touch the yellow ruler and spin the white eraser and push the blue ruler and touch the red pencil

10 Pick up the red pencil then put it into the blue folder and push the yellow ruler and pick up the green eraser then put it into the white bag
11 Spin the blue ruler and touch the red pencil and pick up the green eraser then put it into the yellow bag and push the black pencil
12 Touch the white eraser and push the blue ruler and touch the black pencil and spin the yellow ruler and push the red pencil
13 Pick up the yellow ruler then put it into the white bag and touch the red pencil and push the green eraser and spin the blue ruler
14 Spin the blue ruler and pick up the green eraser then put it into the black box and push the yellow ruler and touch the red pencil

<u>List 3</u>
1 Push the white eraser and touch the green folder and spin the yellow ruler then pick up the white eraser then put it into the red box and push the yellow bag
2 Push the black pencil and touch the blue ruler and spin the red pencil and push the yellow ruler and pick up the white eraser then put it into the black box
3 Push the green eraser and touch the blue ruler and pick up the black pencil then put it into the green folder and spin the red pencil and touch the yellow ruler
4 Pick up the red pencil then put it into the green folder and push the yellow ruler and pick up the green eraser then put it into the white bag and spin the blue ruler
5 Pick up the yellow ruler then put it into the white bag and touch the red pencil and push the green eraser and spin the blue ruler and touch the black pencil
6 Spin the green eraser and pick up the black pencil then put it into the red box and push the white eraser and touch the yellow ruler and spin the red pencil
7 Pick up the white eraser then put it into the red box and touch the yellow ruler and pick up the black pencil then put it into the white bag and push the blue folder
8 Touch the blue ruler and push the white eraser and touch the black pencil and spin the red pencil and pick up the yellow ruler then put it into the red box
9 Pick up the white eraser then put it into the blue folder and spin the red pencil and push the blue ruler and touch the black pencil and spin the yellow ruler
10 Push the yellow ruler and touch the green eraser and pick up the black pencil then put it into the green folder and push the red pencil and touch the white bag
11 Pick up the black pencil and put it into the white bag and spin the blue ruler and pick up the green eraser then put it into the black box and touch the red pencil
12 Spin the blue ruler and pick up the green eraser then put it into the black box and push the yellow ruler and touch the red pencil and push the white eraser
13 Spin the red pencil and touch the blue ruler and pick up the black pencil then put it into the yellow bag and push the green eraser and spin the yellow ruler
14 Touch the white eraser and pick up the black pencil then put it into the yellow bag and touch the green eraser and pick up the blue ruler then put it into the red box

Appendix 4: Lists of three-digit numbers in Experiment 3 and 6

List 1		List 2		List 3	
Articulatory suppression	Backward counting	Articulatory suppression	Backward counting	Articulatory suppression	Backward counting
397	936	156	185	534	645
198	697	213	257	846	247
252	410	136	414	781	298
854	571	289	712	518	564
538	598	539	537	931	960
326	304	716	513	213	184
975	369	740	819	642	361
671	186	851	491	563	619
492	727	591	523	314	283
189	483	337	671	127	179
328	340	143	324	210	395
912	784	847	609	390	164
125	257	902	821	412	780
278	582	623	145	652	242

Appendix 5: Lists of instructions in Experiment 4 and 5

List 1

1 Push the black pencil and spin the green rubber and pick up the red pencil then put it into the blue folder and touch the white bag
2 Touch the red box and push the yellow ruler and pick up the green rubber then put it into the black box and spin the blue ruler
3 Pick up the green rubber then put it into the white bag and spin the yellow ruler and touch the red pencil and push the blue folder
4 Spin the blue ruler and push the green folder and pick up the black pencil then put it into the yellow bag and touch the white rubber
5 Pick up the white rubber and put it into the green folder then spin the yellow ruler and push the black box and touch the red pencil
6 Touch the white bag and spin the red pencil and push the blue ruler then pick up the black pencil and put it into the blue folder
7 Push the black pencil and pick up the yellow ruler then put it into the white bag and touch the blue folder and spin the green rubber
8 Pick up the white rubber then put it into the black box and spin the blue ruler and push the black pencil and touch the green rubber
9 Spin the yellow ruler and touch the green folder and pick up the white rubber then put it into the red box and push the black pencil
10 Touch the red pencil and push the black box and spin the white rubber and pick up the blue ruler then put it into the yellow bag
11 Pick up the yellow ruler then put it into the white bag and push the blue ruler and spin the white rubber and touch the red box
12 Push the red box and pick up the black pencil then put it into the yellow bag and touch the red pencil and spin the blue ruler
13 Spin the blue ruler and push the green folder and touch the red pencil and pick up the white rubber then put it into the yellow bag
14 Spin the red pencil and pick up the yellow ruler then put it into the blue folder and touch the white bag and push the green rubber

List 2

1 Push the white rubber and touch the green folder and spin the yellow ruler and pick up the green rubber then put it into the red box
2 Pick up the yellow ruler then put it into the green folder and touch the red pencil and push the black box and spin the blue ruler
3 Touch the blue ruler and spin the red pencil and push the yellow bag and pick up the white rubber then put it into the black box
4 Push the yellow bag and touch the blue ruler and pick up the black pencil then put it into the green folder and spin the red pencil
5 Pick up the red pencil then put it into the green folder and push the yellow ruler and spin the white rubber and touch the black box
6 Spin the green rubber and pick up the black pencil then put it into the red box and push the white rubber and touch the blue folder
7 Pick up the white rubber then put it into the red box and push the yellow ruler and spin the black pencil and touch the white bag
8 Push the blue folder and touch the black pencil and spin the white rubber and pick up the yellow ruler then put it into the red box

9 Pick up the white rubber then put it into the blue folder and spin the red pencil and push the blue ruler and touch the black box
10 Push the yellow bag and touch the green rubber and pick up the black pencil then put it into the green folder and spin the white rubber
11 Spin the black pencil and touch the white bag and push the blue ruler and pick up the green rubber then put it into the black box
12 Spin the blue ruler and pick up the green rubber then put it into the red box and push the yellow bag and touch the black pencil
13 Push red pencil and touch blue folder and pick up black pencil then put it into yellow bag and spin green rubber
14 Push the yellow bag and pick up the blue ruler then put it into the red box and touch the green rubber and spin the black pencil

List 3

1 Touch the green folder and spin the yellow ruler and push the white rubber and pick up the red pencil then put it into the black box
2 Push the red pencil and touch the green folder and pick up the yellow ruler then put it into the blue folder and spin the white rubber
3 Push the green rubber and touch the white bag and pick up the black pencil then put it into the green folder and spin the red pencil
4 Push the white rubber and touch the red box and spin the yellow ruler and pick up the black pencil then put it into the white bag
5 Pick up the blue ruler then put it into the yellow bag and spin the black pencil and touch the green rubber and push the red box
6 Touch the green rubber and pick up the red pencil then put it into the white bag and push the black box and spin the blue ruler
7 Pick up the black pencil then put it into the red box and push the white bag and spin the green rubber and touch the yellow ruler
8 Spin the blue ruler and push the black pencil and touch the green folder and pick up the yellow ruler then put it into the red box
9 Pick up the red pencil then put it into the blue folder and spin the white rubber and push the yellow bag and touch the green folder
10 Spin the yellow ruler and touch the white rubber and pick up the red pencil then put it into the green folder and push the black box
11 Pick up the red pencil then put it into the green folder and push the blue ruler and touch the yellow bag and spin the black pencil
12 Spin the yellow ruler and push the red box and pick up the black pencil then put it into the green folder and touch the white rubber
13 Push the green folder and touch the blue ruler and pick up the black pencil then put it into the yellow bag and spin the red pencil
14 Spin the blue ruler and pick up the green rubber then put it into the white bag and push the red pencil and touch the yellow bag

List 4 (the additional list in Experiment 5)
1 Push the yellow bag and touch the black box and pick up the blue ruler then put it into the green folder and spin the red pencil
2 Pick up the white rubber then put it into the red box and push the black pencil and spin the yellow ruler and touch the white bag
3 Spin the yellow ruler and pick up the white rubber then put it into the black box and touch the green rubber and push the blue folder

4 Pick up the black pencil then put it into the green folder and push the blue ruler and touch the yellow bag and spin the black pencil (a mistake, replaced it with trial 2 in statistics)
5 Touch the white bag and pick up the green rubber then put it into the yellow bag and spin the blue ruler and push the red pencil
6 Push the blue ruler and touch the white bag and pick up the yellow ruler then put it into the green folder and push the black pencil
7 Pick up the black pencil then put it into the blue folder and spin the red pencil and push the green folder and touch the white rubber
8 Push white rubber and touch red box and spin yellow ruler and pick up white rubber then put it into green folder
9 Push the black pencil and touch the yellow ruler and pick up the white rubber then put it into the black box and spin the red pencil
10 Pick up the red pencil then put it into the blue folder and push the yellow ruler and spin the green rubber and touch the white bag
11 Spin the blue ruler and touch the red box and pick up the green rubber then put it into the yellow bag and push the black pencil
12. Touch green folder and push black pencil and pick up blue ruler then put it into red box and spin white rubber
13 Pick up the yellow ruler then put it into the white bag and touch the red pencil and push the green folder and spin the blue ruler
14 Spin the blue ruler and pick up the green rubber then put it into the black box and push the yellow ruler and touch the red box

Appendix 6: Lists of instructions in Experiment 6 and 7

Each instructional message contained five actions that presented as separate lines in the centre of a computer screen.
Here is an example:

Push black pencil

Spin green rubber

Pick up red pencil

Put it into blue folder

Touch white bag

List 1

1 Push black pencil Spin green rubber Pick up red pencil Put it into blue folder Touch white bag
2 Touch red box Push yellow ruler Pick up green rubber Put it into black box Spin blue ruler
3 Pick up green rubber Put it into white bag Spin yellow ruler Touch red pencil Push blue folder
4 Spin blue ruler Push green folder Pick up black pencil Put it into yellow bag Touch white rubber
5 Pick up white rubber Put it into green folder Spin yellow ruler Push black box Touch red pencil
6 Touch white bag Spin red pencil Push blue ruler Pick up black pencil Put it into blue folder
7 Push black pencil Pick up yellow ruler Put it into white bag Touch blue folder Spin green rubber
8 Pick up white rubber Put it into black box Spin blue ruler Push black pencil Touch green rubber
9 Spin yellow ruler Touch green folder Pick up white rubber Put it into red box Push black pencil
10 Touch red pencil Push black box Spin white rubber Pick up blue ruler Put it into yellow bag

11 Pick up yellow ruler Put it into white bag Push blue ruler Spin white rubber Touch red box
12 Push red box Pick up black pencil Put it into yellow bag Touch red pencil Spin blue ruler
13 Spin blue ruler Push green folder Touch red pencil Pick up white rubber Put it into yellow bag
14 Spin red pencil Pick up yellow ruler Put it into blue folder Touch white bag Push green rubber

<u>List 2</u>
1 Push white rubber Touch green folder Spin yellow ruler Pick up green rubber Put it into red box
2 Pick up yellow ruler Put it into green folder Touch red pencil Push black box Spin blue ruler
3 Touch blue ruler Spin red pencil Push yellow bag Pick up white rubber Put it into black box
4 Push yellow bag Touch blue ruler Pick up black pencil Put it into green folder Spin red pencil
5 Pick up red pencil Put it into green folder Push yellow ruler Spin white rubber Touch black box
6 Spin green rubber Pick up black pencil Put it into red box Push white rubber Touch blue folder
7 Pick up white rubber Put it into red box Push yellow ruler Spin black pencil Touch white bag
8 Push blue folder Touch black pencil Spin white rubber Pick up yellow ruler Put it into red box
9 Pick up white rubber Put it into blue folder Spin red pencil Push blue ruler Touch black box
10 Push yellow bag Touch green rubber Pick up black pencil Put it into green folder Spin white rubber
11 Spin black pencil Touch white bag Push blue ruler Pick up green rubber Put it into black box
12 Spin blue ruler Pick up green rubber Put it into red box Push yellow bag Touch black pencil
13 Push red pencil Touch blue folder Pick up black pencil Put it into yellow bag Spin green rubber
14 Push yellow bag Pick up blue ruler Put it into red box Touch green rubber Spin black pencil

<u>List 3</u>
1 Touch green folder Spin yellow ruler Push white rubber Pick up red pencil Put it into black box
2 Push red pencil Touch green folder Pick up yellow ruler Put it into blue folder Spin white rubber
3 Push green rubber Touch white bag Pick up black pencil Put it into green folder Spin red pencil
4 Push white rubber Touch red box Spin yellow ruler Pick up black pencil Put it into white bag
5 Pick up blue ruler Put it into yellow bag Spin black pencil Touch green rubber Push red box

6 Touch green rubber Pick up red pencil Put it into white bag Push black box Spin blue ruler
7 Pick up black pencil Put it into red box Push white bag Spin green rubber Touch yellow ruler
8 Spin blue ruler Push black pencil Touch green folder Pick up yellow ruler Put it into red box
9 Pick up red pencil Put it into blue folder Spin white rubber Push yellow bag Touch green folder
10 Spin yellow ruler Touch white rubber Pick up red pencil Put it into green folder Push black box
11 Pick up red pencil Put it into green folder Push blue ruler Touch yellow bag Spin black pencil
12 Spin yellow ruler Push red box Pick up black pencil Put it into green folder Touch white rubber
13 Push green folder Touch blue ruler Pick up black pencil Put it into yellow bag Spin red pencil
14 Spin blue ruler Pick up green rubber Put it into white bag Push red pencil Touch yellow bag

Appendix 7: Strategy questionnaire in Experiment 6 and 7

Gender Male ☐ Female ☐

Age ______________ Department_____________

Please check the strategy (can be multiple choices)

<u>Baseline</u> (used in Experiment 6 and 7)

- ☐ Repeating in word (Rehearse)
- ☐ Remember the words visually
- ☐ Imagining doing it in head
- ☐ Grouping the actions
- ☐ Use acronyms
- ☐ No strategy

If you are using any other strategy, please state

____

<u>Repeating numbers</u> (used in Experiment 6 and 7)

- ☐ Repeating in word (Rehearse)
- ☐ Remember the words visually
- ☐ Imagining doing it in head
- ☐ Grouping the actions
- ☐ Think less about the repeating numbers
- ☐ Use acronyms
- ☐ No strategy

If you are using any other strategy, please state

____

Decrease numbers by 2 (used only in Experiment 6)

- ☐ Repeating in word (Rehearse)
- ☐ Remember the words visually
- ☐ Imagining doing it in head
- ☐ Grouping the actions
- ☐ Think less about the decrease numbers
- ☐ Use acronyms
- ☐ No strategy

If you are using any other strategy, please state

____

Tapping (used only in Experiment 7)

- ☐ Repeating in word (Rehearse)
- ☐ Remember the words visually
- ☐ Imagining doing it in head
- ☐ Grouping the actions
- ☐ Think less about the tapping
- ☐ Use acronyms
- ☐ No strategy

If you are using any other strategy, please state

____

Appendix 8: Three-digit numbers in Experiment 7

Practice for articulatory suppression condition: 625 185

Practice for tapping condition: 639 173

The articulatory suppression condition: 936 697 410 571 598 304 369 186 725
483 340 784 257 582

The tapping condition: 397 189 258 853 326 975 528 691 821 296 481 328 912 278

Appendix 9: Summary of main results in seven experiments

Exp	Modality	Setting	Concurrent task1	Concurrent task 2	Phonological loop	Visuospatial sketchpad	Central executive	Recall	Interaction (concurrent task and recall)
1	Spoken	Computer	Articulatory suppression	Backward counting in 3	0.461**		0.897 **	0.019	0.107
2	Spoken	Computer	Articulatory suppression		0.387 *			0.021	0.014
3	Spoken	3D	Articulatory suppression	Backward counting in 2	0.154 [a]		0.819 **	0.349 *	0.030
4	Spoken	3D	Articulatory suppression	Spatial tapping of keypad	0.247 *	0.139 [b]		0.176 *	0.027
5	Spoken	3D	Eye closure			0.899**		0.783**	0.238 [c]
6	Written	3D	Articulatory suppression	Backward counting in 2	0.344 *		0.721**	0.362 *	0.001
7	Written	3D	Articulatory suppression	Corsi-block Tapping	0.402 **	0.657 **		0.297 *	0.007

Note. Baseline condition was always included. The dependent variable was the serial recall of actions. The numbers stand for the effect sizes, η_p^2 . ** stands for $p < 0.001$, and * stands for $p < 0.05$. *a*, $p = 0.058$; *b*, $p = 0.073$; *c*, $p = 0.091$.

Appendix 10: Summary of findings of elements, binding, and order errors across seven experiments

	Phonological loop						Visuospatial sketchpad			Central executive			Action advantage						
	Spoken-computer		Spoken-3D		Written 3D		Spoken-3D		Written 3D	Spoken-computer	Spoken-3D	Written 3D	Spoken-computer		Spoken-3D			Written 3D	
Task	Articulatory suppression						Tapping	Eye closure	Corsi-tapping	Backward counting			Verbal recall vs Action recall						
Experiment	1	2	3	4	6	7	4	5	7	1	3	6	1	2	3	4	5	6	7
Action	*	*	a.s	*	*	*	a.s.	*	*	*	*	*	n.s.	n.s.	*	*	*	*	*
Movement	*	*	n.s.	*	*	*	a.s.	*	*	*	*	*	n.s.	n.s.	*	*	*	*	*
Object or Shape	*	*	n.s.	*	*	*	*	*	*	*	*	*	*	n.s.	*	*	*	*	n.s.
Colour			n.s.	*	*	*	*	*	*		*	*			*	*	*	*	n.s.
Colourful object			n.s.	*	*	*	*	*	*		*	*			*	*	*	*	n.s.
Bind colour							n.s.	*	*		*	n.s.			n.s.	n.s.	n.s.	n.s.	n.s.
Bind movement							n.s.	n.s.	n.s.	n.s.	n.s.	n.s.			n.s.	a.s.	n.s.	n.s.	n.s.
Order error			n.s.	n.s.	n.s.	n.s.	n.s.	*	n.s.		*	*			n.s.	n.s.	n.s.	n.s.	n.s.

Note. 3D stands for the three-dimensional instructional task. * stands for $p < 0.05$; a.s. stands for approaching significance, $0.05 < p < 0.10$; n.s. stands for non significant, $p > 0.10$. Shaded grid represents the effect that was not investigated.

Appendix 11: Summary of strategies in seven experiments

	Experiments							
	Computer-Spoken		3D-Spoken			3D-Written		
Strategies	1	2	3	4	5	6	7	Average
Visual tracking	61	91	93	52	31	a	a	66
Imagining carrying out the action	17	18	40	48	35	81	82	46
Group actions	11	9	33	5	2	42	55	22
Rehearsal	22	0	7	19	19	29	36	19
Decreasing interference	17	9	40	5	0	24	24	17
Assign acronyms	6	0	0	10	0	0	0	2
Binding elements	0	0	13	0	0	0	0	2
Focus on start and end	11	0	0	0	0	0	3	2
Remember words visually	n.a.	n.a.	n.a.	n.a.	n.a.	22	29	26

Note. In the strategy table in Experiment 1 and 2, visual tracking was named as drawing liens between objects. In Experiment 1-5, participants report their strategies and in Experiment 6 and 7, they were given choices. a. the choice of visual tracking was not provided in Experiment 6 and 7; n.a. stands for non-applicable.

References

Aaronson, D. (1968). Temporal course of perception in an immediate recall task. *Journal of Experimental Psychology, 76*, 129-140.

Aaronson, D., & Scarborough, H. S. (1976). Performance theories of sentence coding: Some quantitative evidence. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 56-70.

Adams, J. A. (1971). A closed-loop theory of motor learning. *Journal of motor behavior, 3*(2), 111-150.

Allen, R. J. (2009). *Do what you're told: Encoding instructions for action in working memory.* Paper presented at the BPS Cognitive Psychology Section Conference, University of Hertfordshire.

Allen, R. J. (2011). The role of visual support in following spoken instructions. Unpublished raw data.

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of Experimental Psychology: General, 135*(2), 298-313.

Allen, R. J., & Gathercole, S. E. (2008). *The role of working memory resources in performing actions*. Paper presented at the Experimental Psychology Society Annual Meeting

Allen, R. J., Hitch, G. J., & Baddeley, A. D. (2009). Cross-modal binding and working memory. *Visual Cognition, 17*(1), 83-102.

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and Visuospatial Short Term and Working Memory in Children: Are They Separable? *Child development, 77*(6), 1698-1716.

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, Now you don't: Mediating the mapping between language and visual world. In J. M.Henderson & F. Ferreira (Eds.), *The interface of language, vision and action: Eye movements and the visual world* (pp. 347-386). New York: Psychology press.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation: advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.

Awh, E., Jonides, J., & Renter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 780-790.

Bäckman, L. (1985). Further evidence for the lack of adult age differences on free recall of subject-performed tasks: The importance of motor action. *Human Learning, 4*(2), 79-87.

Bäckman, L., & Nilsson, L. G. (1984). Aging effects in free recall: An exception to the rule. *Human Learning: Journal of Practical Research & Applications, 3*(1), 53-69.

Baddeley, A. D. (1966). The capacity for generating information by randomization. *Quarterly Journal of experimental psychology, 18*, 119-129.

Baddeley, A. D. (1975). Imagery and visual working memory. In S. Dornic (Ed.), *Attention and performance V.* London: Academic Press.

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

Baddeley, A. D. (1992). Working memory. *Science, 225*, 556-559.

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology, 49A*(1), 5-28.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? . *Trends in Cognitive Science, 4*(11), 417-423.

Baddeley, A. D. (2003a). Working memory and language: an overview. *Journal of communication disorders, 36*(3), 189-208.

Baddeley, A. D. (2003b). Working memory: Looking back and looking forward. *Nature, 4*, 829-839.

Baddeley, A. D. (2007). *Working memory, thought, and action*. New York: Oxford University Press.

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2010). Investigating the episodic buffer. *Psychologica Belgica, 50*, 223-243.

Baddeley, A. D., Allen, R. J., & Hitch, G. J. (2011). Binding in visual working memory: The role of the episodic buffer. *Neuropsychologia*.

Baddeley, A. D., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General, 130*(4), 641.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G.H.Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York Academic Press.

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language*, *61*, 438-456.

Baddeley, A. D., & Larsen, J. D. (2007). The phonological loop unmasked? A comment on the evidence for a "perceptual-gestural" alternative. *Quarterly Journal of Experimental Psychology, 60*(4), 497-504.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, *36*(2), 233-252.

Baddeley, A. D., & Logie, H. L. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York: Cambridge University Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word Length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14* (6), 575-589.

Baggett, P. (1984). Role of temporal overlap of visual and auditory material in forming dual media associations. *Journal of Educational Psychology, 76*, 408-417.

Baggett, P., & Ehrenfeucht, A. (1983). Encoding and retaining information in the visuals and verbals of an educational movie. *Educational Communications and Technology Journal*, *31*, 23-32.

Baker, S. C., Frith, C. D., Frackowiak, R. S., & Dolan, R. J. (1996). Active representation of shape and spatial location in man. *Cerebral Cortex 6*(4), 612-619.

Barnat, S. B., Klein, P. J., & Meltzoff, A. N. (1996). Deferred imitation across changes in context and object: Memory and generalization in 14-month-old infants. *Infant Behavior and Development, 19*, 241-251.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 570.

Bauer, P. J., & Mandler, J. (1989). One thing follows another: Effects of temporal structure on 1- and 2- year - olds' recall of events. *Developmental Psychology, 25*, 197-206.

Bauer, P. J., & Shore, C. M. (1987). Making a memorable event: Effects of familiarity and organization on young children's recall of action sequence. *Cognitive Development, 2*(4), 327-338.

Bekkering, H., Wohlschläger, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *Quarterly Journal of experimental psychology, 8*, 319-341.

Berch, D. B., Krikorian, R., & Huha, E. M. (1998). The Corsi block-tapping task: methodological and theoretical considerations. *Brain and Cognition, 38*, 317-338.

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology, 26*, 467-482.

Brockmole, J. R. (2009). *The visual world in memory*. Hove, England: Psychology Press.

Brooks, L. R. (1967). The suppression of visualisation by reading. *Quarterly Journal of Experimental Psychology, 19*.

Brown, L. A., & Brockmole, J. R. (2010). The role of attention in binding visual features in working memory: Evidence from cognitive ageing. *Quarterly Journal of Experimental Psychology, 63*(10), 2067-2079.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience 13*, 400-404.

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31*(4), 429-460.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*(3), 551.

Burnken, R., Plass, J., L., & Moreno, R. (2010). Current issues and open questions in cognitive load research. In J. Plass, L., R. Moreno & R. Brunken (Eds.), *Cognitive load theory* (pp. 253-272). New York: Cambridge University Press.

Busch, R. M., Farrell, K., Lisdahl-Medina, K., & Krikorian, R. (2005). Corsi block-tapping task performance as a function of path configuration. *Journal of Clinical and Experimental Neuropsychology, 27*(1), 127-134.

Butcher, K. R. (2006). Learning from text and diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology, 98,* 182-197.

Byrne, R. W. (2002). Seeing actions as hierarchically organized structures: Great ape manual skills. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind* (pp. 122-142). Cambridge: Cambridge University Press.

Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences, 21*, 667-721.

Carey, D. P. (1996). Neurophysiology: 'Monkey see, monkey do' cells. *Current Biology, 6*(9), 1087-1088.

Chan, R. C. K., Shum, D., Toulopoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology, 23*, 201–216.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293-332.

Chartand, T. L., & Barge, J. A. (1999). The chameleon effect: the preception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*, 893-910.

Clark, H. H., & Clark, E. V. (1968). Semantic distinctions and memory for complex sentences. *Quarterly Journal of Experimental Psychology, 20*, 129-138.

Clark, R. C., & Mayer, R. (2008). *E-learning and the science of instruction*. San Francisco: John Wiley & Sons, Inc.

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research, 53*, 445-459.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42*(2), 21-30.

Clark, R. E., & Salomon, G. (1986). Media in teaching. In M.C.Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 464-478). New York: Macmillan.

Cohen, R. L. (1981). On the generality of some memory laws. *Scandinavian Journal of Psychology, 22*(1), 267-281.

Cole, K. N., & Fewell, R. R. (1983). A quick language screening test for young children: The Token Test. *Journal of Psychoeducational Assessment, 1*(2), 149-153.

Coluccia, E., Bosco, A., & Brandimonte, M. (2007). The role of visuo-spatial working memory in map learning: new findings from a map drawing paradigm. *Psychological Research, 71*, 359–372.

Conrad, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology, 55*, 75-84.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55,* 429-432.

Cooper, G., Tindall-Ford, S., Chandler, P., & Sweller, J. (2001). Learning by imagining. *Journal of Experimental Psychology: Applied 7*, 68-82.

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34*(2-B), 819.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). New York: Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87-114.

Cowan, N. (2005). *Working memory capacity*. New York: Psychology Press.

Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language, 46*(1), 153-177.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia learning environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94*, 428-434.

Craighero, L., Fadiga, L., Rizzolatti, G., & Umiltà, C. (1998). Visuomotor priming. *Mechanisms of Visual Attention: A Cognitive Neuroscience Perspective*, 109-126.

Craighero, L., Fadiga, L., Umiltà, C. A., & Rizzolatti, G. (1996). Evidence for visuomotor priming effect. *Neuroreport, 8*(1), 347.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Craik, F. I. M., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 12,* 599-607.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*(4), 422-433.

De Renzi, E., & Faglioni, P. (1978). Normative data and screening power of a shortened version of the Token Test. *Cortex, 14,* 41-49.

De Renzi, E., & Vignolo, L. (1962). The Token test: A sensitive test to detect receptive disturbances in aphasics. *Brain, 85*(4), 665-678.

Decety, J. (2002). Is there such as a thing as functional equivalence between imagined, observed, and executed action? In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind* (pp. 291-310). Cambridge: Cambridge University Press.

Decety, J., & Ingvar, D. H. (1990). Brain structures participating in mental simulation of motor behavior: A neuropsychological interpretation. *Acta Psychologica, 73*, 13-34.

Decety, J., Jeannerod, M., & Prablanc, C. (1989). The timing of mentally represented actions. *Behavioral and Brain Research, 34*, 35-42.

Decety, J., & Michel, F. (1989). Comparative analysis of actual and mental movement times in two graphic tasks. *Brain and Cognition, 11*(1), 87-97.

Della Sala, S., Gray, C., Baddeley, A. D., Allemano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neuropsychologia, 37,* 1189-1199.

Della Sala, S., & Logie, R. H. (2002). Neuropsychological impairments of visual and spatial working memory. *Handbook of memory disorders*, 271-292.

Depoorter, A., & Vandierendonck, A. (2009). Evidence for modality-independent order coding in working memory. *The Quarterly Journal of Experimental Psychology, 62*(3), 531-549.

Deyzac, E., Logie, R. H., & Denis, M. (2006). Visuospatial working memory and the processing of spatial descriptions. *British Journal of Psychology, 97,* 217–243.

Driskell, J., Copper, C., & Moran, A. (1994). Does mental practice enhance performance? *Journal of Applied Psychology, 79*(4), 481-492.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology-General, 123*(2), 161-176.

Engelkamp, J. (2001). Action memory: A system-oriented approach. In H. D. Zimmer, R.Cohen, M. Guynn, J. Engelkamp, R. Kormi-Nouri & M. Foley (Eds.), *Memory for Action: A Distinct Form of Episodic Memory*. New York: Oxford University Press.

Engelkamp, J., & Zimmer, H. D. (1985). Motor programs and their relation to semantic memory. *German Journal of psychology*.

Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual Differences in working memory for comprehension and following directions. *Journal of Educational Research, 84*(5), 253-262.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102-134). New York: Cambridge University Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 125*, 309-331.

Eom, W., & Reister, R. A. (2000). The effects of self-regulation and instructional control on performance and motivation in computer-based instruction. *International Journal of Instructional Media, 27*(3), 247-260.

Ericsson, K. A., & Delaney, P. (1999). Long term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 257-297). New York: Cambridge University Press.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. . *Psychological Review, 102*(2), 211-245.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor Facilitation During Action Observation: A Magnetic Stimulation Study. *Journal of Neurophysiology, 73*, 2608-2611.

Farah, M. J., Hammond, K. M., Levine, D. N., & Calvanio, R. (1988). Visual and spatial mental imagery: Dissociable systems of representation. *Cognitive Psychology*, *20*, 439-462.

Farmer, E. W., Berman, J. V. F., & Fletcher, Y. L. (1986). Evidence for a visuospatial scratch-pad in working-memory. *Quarterly Journal of Experimental Psychology*, *38A*, 675-688.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9*(1), 59-79.

Fisk, J. E., & Sharp, C. A. (2003). The role of the executive system in visuo-spatial memory functioning. *Brain and Cognition, 52*(3), 364-381.

Flanagan, J. R., and Johansson, R.S. (2003). Action plans used in action observation. *Nature, 424*, 769-771.

Fleming, M., & Levie, W. H. (1993). *Instructional message design: Principles from the behavioural and cognitive sciences*. Englewood Cliffs, NJ: Educational Technology Publications.

Fletcher, J. D., & Tobis, S. (2005). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 117-134). New York: Cambridge University Press.

Frisk, J. E., & Sharp, C. A. (2003). The role of the executive system in visuo-spatial memory functioning. *Brain and Cognition, 52*(3), 364-381.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119*, 593-609.

Garden, S., Cornoldi, C., & Logie, R. H. (2002). Visuo-spatial working memory in navigation. *Applied Cognitive Psychology, 16*, 35-50.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*(4), 513-543.

Gathercole, S. E., & Alloway, T. P. (2008). *Working memory and learning: A guide for teachers.*: Sage Publishing.

Gathercole, S. E., Durling, E., Evans, M., Jeffcock, S., & Stone, S. (2008). Working memory abilities and children's performance in laboratory analogues of classroom activities. *Applied Cognitive Psychology, 22*.

Gattis, M., Bekkering, H., & Wohlschläger, A. (2002). Goal-directed imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind* (pp. 183-205). Cambridge: Cambridge University Press.

Gibson, J. J. (1977). The theory of affordances. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67-82). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and instruction, 15*, 313-331.

Ginns, P., Chandler, P., & Sweller, J. (2003). When imagining information is effective. *Contemporary Educational Psychology, 28*, 229-251.

Glenberg, A. M. (1997). What memory is for. *Behavioural & brain sciences, 20*(1), 55.

Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(1), 3.

Goodale, M., Jakobson, L., & Keillor, J. (1994). Differences in the visual control of pantomimed and natural grasping movements. *Neuropsychologia, 32*(10), 1159-1178.

Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A Meta-analysis. *Human Brain Mapping, 12*, 1-19.

Grafton, S. T., Fadiga, L., Arbib, M. A., & Rizzolatti, G. (1997). Premotor cortex activation during observation and naming of familiar tools. *Neuroimage 6*, 231-236.

Griffin, Z. M. (2004). Why look? Reasons for eye movements related to language production. In J. M.Henderson & F. Ferreira (Eds.), *The interface of language, vision and action: Eye movements and the visual world* (pp. 213-248). New York: Psychology press.

Hale, S., Myerson, J., Rhee, S. H., Weiss, C. S., & Abrams, R. A. (1996). Selective interference with the maintenance of location information in working memory. *Neuropsychology, 10*(2), 228.

Hall, D., & Gathercole, S. E. (2011). Serial recall of rhythms and verbal sequences: Impacts of concurrent tasks and irrelevant sound. *Quarterly Journal of Experimental Psychology*, 1.

Hanley, J. R., & Broadbent, C. (1987). The effect of unattended speech on serial recall following auditory presentation. *British Journal of Psychology, 78*, 287-297.

Hanna, E., & Meltzoff, A. N. (1993). Peer imitation by toddlers in laboratory, home and day-care contexts: Implications for social learning and memory. *Developmental Psychology, 29*, 701-710.

Harp, S. F., & Mayer, R. E. (1998). How the seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*, 414-434.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*, 301–307.

Hautzel, H., Mottaghy, F. M., Schmidt, D., Zemb, M., Shah, N. J., Müller-Gärtner, H.-W., et al. (2002). Topographic segregation and convergence of verbal, object, shape and spatial working memory in humans. *Neuroscience Letters, 323*(2), 156-160.

Henderson, J. M., & Ferreira, F. (2004). *The interface of language, vision and action: Eye movements and the visual world*. New York: Psychology press.

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36*(2), 73-137.

Henson, R. N. A., Burgess, N., & Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia, 38*(4), 426-440.

Horz, H., & Schnotz, W. (2010). Cognitive load in learning with multiple representations. In J. Plass, L., R. Moreno & R. Brunken (Eds.), *Cognitive load theory* (pp. 229-252). New York: Cambridge University Press.

Hurlstone, M. J. (2010). *The problem of serial order in visuospatial short-term memory.* University of York, York,UK.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*(2526-2528).

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia, 33*(11), 1419-1432.

Jeannerod, M. (1997a). Action planning *The Cognitive Neuroscience of Action* (pp. 126-151). Oxford, UK: Blackwell Publishers Inc.

Jeannerod, M. (1997b). *The cognitive neuroscience of action*. Oxford, UK: Blackwell Publishers Inc.

Jeannerod, M. (1997c). The contribution of mental imagery to understanding motor representations *The Cognitive Neuroscience of Action* (pp. 94-125). Oxford, UK: Blackwell Publishers Inc.

Jeannerod, M. (1997d). Design for a motor representation *The Cognitive Neuroscience of Action* (pp. 164-191). Oxford, UK: Blackwell Publishers Inc.

Jeannerod, M. (1997e). Neural substrates for object-oriented actions *The Cognitive Neuroscience of Action* (pp. 11-55). Oxford, UK: Blackwell Publishers Inc.

Jefferies, E., Ralph, M. A. L., & Baddeley, A. D. (2004). Automatic and controlled processing in sentence recall: The role of long-term and working memory. *Journal of Memory and Language, 51,* 623-643.

Jenkins, I. H., Brooks, D. J., Nixon, P. D., Frackowiak, R. S., & Passingham, R. E. (1994). Motor sequence learning: a study with positron emission tomography. *The Journal of Neuroscience, 14*(6), 3775-3790.

Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995a). The functional equivalence of verbal and spatial information in serial short-term memory. *journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 108-118.

Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995b). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 1008.

Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science, 38*.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329.

Kalyuga, S. (2010). Schema acquisition and sources of cognitive load. In J. Plass, L., R. Moreno & R. Brunken (Eds.), *Cognitive load theory* (pp. 109-130). New York: Cambridge University Press.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist 38,* 23-31.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human factors, 40,* 1-17.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13,* 351-372.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*, 126-136.

Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology, 21*, 5-23.

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human factors, 46*, 567-581.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579-588.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology-General, 133*(2), 189-217.

Kaplan, C., & White, M. (1980). Children's Direction-following behavior in Grades K-5. *The Journal of Educational Research, 74*(1), 43-48.

Karlsen, P. J., Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2010). Binding across space and time in visual working memory. *Memory and Cognition, 38*, 292-303.

Keele, S. W., Ivry, R., Mayr, U., & Hazeltine, E. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review, 110*(2), 316-339.

Kester, L., Paas, F., & van Merrienboer, J. J. G. (2010). Instructional control of cognitive load in the design of complex learning environment. In J. Plass, L., R. Moreno & R. Brunken (Eds.), *Cognitive load theory* (pp. 109-130). New York: Cambridge University Press.

Kim, E. S., Bayles, K. A., & Beeson, P. M. (2008). Instruction processing in young and older adults: Contributions of memory span. *Aphasiology, 22*(7), 753-762.

Klatzky, R., MacWhinney, B., & Behrmann, M. (2008). *Embodiment, Ego-space, and Action*. New York: Psychology Press.

Klauer, K., & Zhao, Z. (2004). Double Dissociations in Visual and Spatial Short-Term Memory. *Journal of Experimental Psychology: General, 133*(3), 355-381.

Klauer, K. C., & Stegmaier, R. (1997). Interference in immediate spatial memory: Shifts of spatial attention or central-executive involvement? *Quarterly Journal of Experimental Psychology, 50A*(1), 79-99.

Klein, P. J., & Meltzoff, A. N. (1999). Long-term memory, forgetting, and deferred imitation in 12-month-old infants. *Developmental Science, 2*, 102-113.

Kohl, R. M., & Roenker, D. L. (1980). Bilateral transfer as a function of mental imagery. *Journal of motor behavior, 12*, 197-206.

Kohl, R. M., & Roenker, D. L. (1983). Mechanism involvement during skill imagery. *Journal of motor behavior, 15*(2), 179-190.

Koriat, A., Ben-Zur, H., & Nussbaum, A. (1990). Encoding information for future action: Memory for to-be-performed tasks versus memory for to-be-recalled tasks. *Memory & Cognition, 18*(6), 568-578.

Kormi-Nouri, R. (1994). *Memory for action events: an episodic integration view.* Umea university, Umea.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision research, 41*(25-26), 3559-3565.

Larsen, J. D., & Baddeley, A. D. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: Do they have a common source? *Quarterly Journal of Experimental Psychology, 56*(8), 1249-1268.

Lawrence, B. M., Myerson, J., & Abrams, R. A. (2004). Interference with spatial working memory: An eye movement is more than a shift of attention. *Psychonomic Bulletin & Review, 11*(3), 488-494.

Lawrence, B. M., Myerson, J., Oonk, H. M., & Abrams, R. A. (2001). The effects of eye and limb movements on working memory. *Memory, 9*(9), 433-444.

Leahy, W., & Sweller, J. (2004). Cognitive load and the imagination effect. *Applied Cognitive Psychology, 18*, 857-875.

Leahy, W., & Sweller, J. (2005). Interactions among the imagination, expertise reversal and element interactivity effects *Journal of Experimental Psychology: Applied 11*, 266-276.

Lesser, R. (1976). Verbal and non-verbal memory components in the token test. *Neuropsychologia, 14*, 79-85.

Lewandowsky, S., & Oberauer, K. (2009). No evidence for temporal decay in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1545-1551.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1-36.

Logie, R. H. (1986). Visuo-spatial processing in working memory. *The Quarterly Journal of Experimental Psychology, 38*(2), 229-247.

Logie, R. H. (1995). *Visuo-spatial Working memory*. East Sussex, UK: Lawrence Erlbaum Associates Ltd.

Logie, R. H. (2003). Spatial and visual working memory: A mental workspace. *Psychology of Learning and Motivation, 42*, 37-78.

Logie, R. H. (2011). The functional organization and capacity limits of working memory. *Current Directions in Psychological Science*.

Logie, R. H., Della Sala, S. D., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition, 24*(3), 305-321.

Logie, R. H., Sala, S. D., Wynn, V., & Baddeley, A. D. (2000). Visual similarity effects in immediate verbal serial recall. *Quarterly Journal of Experimental Psychology, 53*(3), 626-646.

Longoni, A., Richardson, J. T. E., & Aiello, A. (1993). Articulatory rehearsal and phonological storage in working memory. *Memory & Cognition*, *21*(1), 11-22.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281.

Luzzatti, C., Vecchi, T., Agazzi, D., Cesa-Bianchi, M., , & Vergani, C. (1998). A neurological dissociation between visual and spatial processing in mental imagery. *Cortex, 34* 461-469.

Marschark, M., Richman, C. L., Yuille, J. C., & Hunt, R. R. (1987). The role of imagery in memory: On shared and distinctive information. *Psychological Bulletin, 102*(1), 28.

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science, 270*, 102-105.

Masson, M. E. J. (1983). Conceptual processing of text during skimming and rapid sequential reading. *Memory & Cognition, 11*(3), 262-274.

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology, 81*, 240-246.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1-19.

Mayer, R. E. (2005). Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 147-158). New York: Cambridge University Press.

Mayer, R. E., & Anderson, R. (1991). Animation need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83*, 484-490.

Mayer, R. E., & Anderson, R. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, *84*, 64-73.

Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology, 88*, 64-73.

Mayer, R. E., Dow, G., & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds. *Journal of Educational Psychology*, *95*, 806-813.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 88*, 64-73.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*, 187-198.

Mayer, R. E., & Jackson, J. (2005). The case for coherence in scientific explanations: Quantitative details can hurt qualitative understanding. *Journal of Experimental Psychology: Applied, 11*, 13-18.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual coding hypothesis. *Journal of Educational Psychology, 94*(1), 484-490.

Mayer, R. E., & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. Plass, L., R. Moreno & R. Brunken (Eds.), *Cognitive load theory* (pp. 131-152). New York: Cambridge University Press.

Mayer, R. E., Moreno, R., Boire, M., & Vagge, S. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology, 91*, 638-643.

Mayer, R. E., & Sims, V. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389-401.

Mayer, R. E., Steinhoff, K., Bower, G. H., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development, 43*, 31-43.

McCarthy, G. (1996). Activation of human prefrontal cortex during spatial and nonspatial working memory tasks as measured by functional MRI. *Cerebral Cortex, 6*, 600-611.

Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: Long term memory for novel acts and multiple stimuli. *Developmental Psychology, 24*, 470-476.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838-850.

Meltzoff, A. N. (2002). Elements of a developmental theory of imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind* (pp. 1-19). Cambridge: Cambridge University Press.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*, 75–78.

Meltzoff, A. N., & Moore, M. K. (1983). Newborn infants imitate adult facial gestures. *Child development, 54*, 702-709.

Meltzoff, A. N., & Moore, M. K. (1989). Imitation in new born infants: Exploring the range of gestures imitated and the underlying mechanisms. *Developmental Psychology, 25*, 957-962.

Meltzoff, A. N., & Moore, M. K. (1994). Imitation, memory, and representation of persons. *Infant Behavior and Development, 17*, 83–99.

Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development and Parenting, 7*, 179–192.

Meltzoff, A. N., & Moore, M. K. (2002). Imitation, memory, and representation of persons. *Infant Behavior and Development, 25*, 39-61.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition, 66*(2), B25-B33.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*: Holt.

Milner, A. D., & Goodale, M. A. (2006a). Attention, consciousness, and the coordination of behavior. In A. D. Milner & M. A. Goodale (Eds.), *The visual brain in action* (pp. 181-206).

Milner, A. D., & Goodale, M. A. (2006b). Dissociations between perception and action. In A. D. Milner & M. A. Goodale (Eds.), *The visual brain in action* (pp. 157-180).

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British medical bulletin, 27*(3), 272.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in neurosciences, 6,* 414-417.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visualspatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130*(4), 621-640.

Moreno, R., & Mayer, R. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology, 94,* 598-610.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91,* 358-368.

Moreno, R., & Mayer, R. E. (2000a). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92,* 117-125.

Moreno, R., & Mayer, R. E. (2000b). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology, 93*, 724-733.

Moreno, R., Mayer, R. E., Spires, H., & Lester, J. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction, 196*(177-214).

Moreno, R., & Park, B. (2010). Cognitive load theory: Historical Development and Relation. In J. Plass, L., R. Moreno & R. Brunken (Eds.), *Cognitive load theory* (pp. 9-28). New York: Cambridge University Press.

Mousavi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*, 319-334.

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology, 78*, 679-684.

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition, 16*(4), 343-352.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*(3), 251-269.

Neath, I., & Nairne, J. S. (1995). Word-length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review, 2*(4), 429-441.

Neath, I., Surprenant, A. M., & LeCompte, D. C. (1998). Irrelevant speech eliminates the word length effect. *Memory & Cognition, 26*(2), 343-354.

Norman, D. A., & Shallice, T. (1983). Attention to action-willed and automatic control of behavior. *Bulletin of the Psychonomic Society, 21*(5), 354-354.

Norman, D. A., & Shallice, T. (1986). Consciousness and Self-regulation. In R. J. Davidson, G. E. Schwarts & D. Shapiro (Eds.), *Advances in Research and Theory* (pp. 1-18). New York: Plenum.

O'Neil, H. F., Mayer, R. E., Herl, H. E., Niemi, C., Olin, K., & Thurman, R. A. (2000). Instructional strategies for virtual aviation training environments. In H. F. O'Neil & D. H. Andrews (Eds.), *Aircrew training and assessment* (pp. 105-130). Mahwah, NJ: Lawrence Erlbaum.

Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language, 49*(4), 469-483.

Paas, F. G., & van Merrienboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*, 351-371.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761-781.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Wilston.

Paivio, A. (1986a). Dual coding theory *Mental Representation: A Dual Coding Theory* (pp. 53-83). New York: Oxford University Press.

Paivio, A. (1986b). *Mental representation: A dual-coding approach*. Oxford: Oxford University Press.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Paivio, A. (2006). *Dual coding theory and education*. Paper presented at the Pathways to literacy achievement for high poverty children.

Paquier, P. F., Mourik, M., Dongen, H., Catsman-Berrevoets, C., Creten, W. L., & Borsel, J. (2009). Normative data of 300 Dutch-speaking children on the Token Test. *Aphasiology, 23* (4), 427–437.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and aging, 17*(2), 299-320.

Parkhurst, P. E., & Dwyer, F. M. (1983). An experimental assessment of students' IQ level and their ability to profit from visualized instruction *Journal of Instructional Psychology, 10*, 9-20.

Parsons, L. (1994). Temporal and kinematic properties of motor behavior reflected in mentally simulated action. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 709-730.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature, 362*, 342-345.

Pearson, D. G., Logie, R. H., & Gilhooly, K. J. (1999). Verbal representations and spatial manipulation during mental synthesis. *European Journal of Cognitive Psychology, 11*, 295–314.

Phillips, W. A., & Christie, D. F. M. (1977). Interference with visualization *Quarterly Journal of Experimental Psychology, 29*(4), 637-650.

Pickering, S. J., & Gathercole, S. E. (2001). Working Memory Test Battery for Children: Psychological Corporation Europe.

Plass, J., L., Kalyuga, S., & Leutner, D. (2010). Individual difference and cognitive load theory. In J. Plass, L., R. Moreno & R. Brunken (Eds.). New York: Cambridge University Press.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and instruction, 12*, 61-86.

Portrat, S., Barrouillet, P., & Camos, V. (2008). Time-related decay or interference-based forgetting in working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1561.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology, 79*(1p2), 1-16.

Posner, M. I., & Keele, S. W. (1967). Decay of visual information from a single letter. *Science, 158*(3797), 137.

Postle, B. R., Idzikowski, C., Della Sala, S., Logie, R. H., & Baddeley, A. D. (2006). The selective disruption of spatial working memory by eye movements. *Quarterly Journal of experimental psychology, 59*, 100-120.

Postma, A., & De Haan, E. H. F. (1996). What was where? Memory for object locations. *Quarterly Journal of Experimental Psychology, 49*(1), 178-199.

Potter, M. C. (1975). Meaning in visual search. *Science, 187*(4180), 965-966.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology, 9*, 129-154.

Prinz, W. (2002). Experimental approaches to imitation. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind* (pp. 143-162). Cambridge: Cambridge University Press.

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience 17*(6), 884–892.

Purves, D., Brannon, E., Cabeza, R., Huettel, S. A., LaBar, K., Platt, M., et al. (2008). *Principles of Cognitive Neuroscience* Sinauer Associates.

Quinn, J., & McConnell, J. (2006). The interval for interference in conscious visual imagery. *Memory, 14*(2), 241-252.

Quinn, J. G., & Ralston, G. E. (1986). Movement and attention in visual working memory. *Quarterly Journal of Experimental Psychology, 38A*, 689–703.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (2002). From mirror neurons to imitation: Facts and speculations. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind* (pp. 247-266). Cambridge: Cambridge University Press.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor activations. *Cognitive Brain Research, 3*, 131-141.

Rizzolatti, G., Riggio, L., & Sheliga, B. M. (1994). Space and selective attention. In C. U. M. Moscovitch (Ed.), *Attention and Performance* (pp. 231-265). Cambridge, MA: MIT Press.

Rothi, L. J. G., Ochipa, C., & Heilman, K. M. (1991). A cognitive neuropsychological model of limb praxis. *Cognitive Neuropsychology, 8*, 443-458.

Rudkin, S., Pearson, D. G., & Logie, R. H. (2007). Executive processes in visual and spatial working memory tasks. *Quarterly Journal of Experimental Psychology, 60*, 79-100.

Rumiati, R. I., & Tessari, A. (2002). Imitation of novel and well-known actions: The role of short-term memory. *Experimental Brain Research, 142*.

Rumiati, R. I., Weiss, P. H., Tessari, A., Assmus, A., Zilles, K., Herzog, H., et al. (2005). Common and differential neural mechanisms supporting imitation of meaningful and meaningless actions. *Journal of Cognitive Neuroscience 17*(9), 1420-1431.

Saito, S. (1994). What effect can rhythmic finger tapping have on the phonological similarity effect? *Memory & Cognition, 22*(2), 181-187.

Salamé, P., & Baddeley, A. D. (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society, 24*(4), 263-265.

Salomon, G. (1979/1994). *Interaction of media, cognition, and learning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Saltz, E., & Dixon, D. (1982). Let's pretend: The role of motoric imagery in memory for sentences and words* 1. *Journal of Experimental Child Psychology, 34*(1), 77-92.

Salway, A. F. S., & Logie, R. H. (1995). Visuospatial working memory, movement control and executive demands. *British Journal of Psychology, 86*, 253–269.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review, 82*(4), 225.

Seitz, K., & Schumann-Hengsteler, R. (2002). Phonological loop and central executive processes in mental addition and multiplication. *Psychologische Beitrage, 44*(2), 275-302.

Seitz, R. J., & Roland, P. E. (1992). Learning of sequential finger movements in man: a combined kinematic and positron emission tomography (PET) study. *European Journal of Neuroscience, 4*(2), 154-165.

Seymour, P. H. K. (1974). Generation of a pictorial code. *Memory and Cognition, 2*, 224-232.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: a neuropsychological study. *Quarterly Journal of Experimental Psychology, 22*, 261-273.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience, 18*, 555-586.

Smirni, P., Villardita, C., & Zappalà, G. (1983). Influence of different paths on spatial memory performance in the Block-Tapping Test. *Journal of Clinical Neuropsychology, 5*(4), 355-359.

Smith, E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*(12), 1627-1661.

Smith, E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 6*, 11-20.

Smith, E. E., Jonides, J., Koeppe, R. A., Awh, E., Schumacher, E. H., & Minoshima, S. (1995). Spatial versus object working memory: PET investigations. *Journal of Cognitive Neuroscience, 7*, 337-356.

Smyth, M. M., Pearson, N. A., & Pendleton, L., R. (1988). Movement and working memory: Patterns and positions in space. *Quarterly Journal of Experimental Psychology, 40* (3), 497-514.

Smyth, M. M., & Pendleton, L. R. (1989). Working memory for movements. *Quarterly Journal of Experimental Psychology, 41*(2), 235-250.

Smyth, M. M., & Pendleton, L. R. (1990). Space and movement in working memory. *Quarterly Journal of Experimental Psychology, 42*(2), 291-304.

Smyth, M. M., & Scholey, K. A. (1992). Determining spatial span: The role of movement time and articulation rate. *Quarterly Journal of Experimental Psychology, 45*, 479-501.

Smyth, M. M., & Scholey, K. A. (1994). Interference in immediate spatial memory. *Memory & Cognition, 22*, 1-33.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 174-215.

Spivey, M., Richardson, D. C., & Fitneva, S. A. (2004). Thinking outside the brain: Spatial indices to visual and linguistic information. In J. M.Henderson & F. Ferreira (Eds.), *The interface of language, vision and action: Eye movements and the visual world* (pp. 161-190). New York: Psychology press.

Spivey, M. J., & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research, 65*(4), 235-241.

Strafella, A. P., & Paus, T. (2000). Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study *Neuroreport 11*, 2289-2292.

Stuss, D. T., Shallice, T., Alexander, M. P., & Picton, T. W. (1995a). A multidisciplinary approach to anterior attentional functions. *Annals of the New York Academy of Sciences, 769*(1), 191-211.

Stuss, D. T., Shallice, T., Alexander, M. P., & Picton, T. W. (1995b). A multidisciplinary approach to anterior attentional functions. *Annals of the New York Academy of Sciences, 769*, 191-211.

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science, 32*.

Sweller, J. (2010). Cognitive load theory: Recent theoretic advances. In J. Plass, L., R. Moreno & R. Brunken (Eds.), *Cognitive load theory* (pp. 29-47). New York: Cambridge University Press.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185-233.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General*, *119*, 176-192.

Tanenhaus, M. K., Chambers, C. G., & Hanna, J. E. (2004). Referential domains in spoken language comprehension: Using eye movements to bridge the product and action traditions. In J. M.Henderson & F. Ferreira (Eds.), *The interface of language, vision and action: Eye movements and the visual world* (pp. 279-318). New York: Psychology press.

Tessari, A., Canessa, N., Ukmar, M., & Rumiati, R. I. (2007). Neuropsychological evidence for a strategic control of multiple routes in imitation. *Brain, 103*(4), 1111-1126.

Tessari, A., & Rumiati, R. I. (2004). The strategic control of multiple routes in imitation of actions. *Journal of Experimental Psychology: Human Perception and Performance, 30*(6), 1107-1116.

Tindall-Ford, S., & Sweller, J. (2006). Altering the modality of instructions to facilitate imagination. *Instructional Science*, *34*, 343-365.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Culture learning. *Behavioural and brain sciences, 16*, 495-552.

Travis, L. L. (1997). Goal-based organization of event memory in toddlers. In P. W. van den Broek, P. J. Bauer & T. Bourg (Eds.), *Developmental spans in event comprehension of representation: Bridging fictional and actual events* (pp. 111-138). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Tresch, M. C., Sinnamon, H. M., & Seamon, J. G. (1993). Double dissociation of spatial and object visual memory: Evidence from selective interference in intact human subjects. *Neuropsychologia, 31*(3), 211-219.

Tuovinen, J., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*, 334-341.

Ueno, T., Allen, R. J., Baddeley, A. D., Hitch, G. J., & Saito, S. (2011). Disruption of visual feature binding in working memory. *Memory and Cognition, 39*, 12-23.

Ueno, T., Mate, J., Allen, R. J., Hitch, G. J., & Baddeley, A. D. (2011). What goes through the gate? Exploring interference with visual feature binding. *Neuropsychologia*.

Vallar, G., & Papagno, C. (2002). Neuropsychological impairment of verbal short-term memory. In A. D. Baddeley, M. D. Kopelman & B. A. Wilson (Eds.), *Handbook of Memory Disorders* (pp. 249-270). Chichester: Wiley.

van Leeuwen, M. L., Baaren, R. B., Martin, D., Dijksterhuisa, A., & Bekkering, H. (2009). Executive functioning and imitation: Increasing working memory load facilitates behavioural imitation. *Neuropsychologia, 47*, 3265-3270.

Vandierendonck, A., De Vooght, G., & van der Goten, K. V. (1998). Interfering with the central executive by means of a random interval repetition task. *Quarterly Journal of Experimental Psychology, 51A*, 197-218.

Vandierendonck, A., Kemps, E., Fastame, M., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology, 95*, 57-79.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 92-114.

Vogt, S. (2002). Visuomotor couplings in object-oriented and imitative actions. In A. N. Meltzoff & W. Prinz (Eds.), *The imitative mind* (pp. 206-220). Cambridge: Cambridge University Press.

Wager, T. D., & Smith, E. E. (2003a). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective, & Behavior Neuroscience, 3*(4), 255-274.

Wager, T. D., & Smith, E. E. (2003b). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience, 3*(4), 255-274.

Warrington, E. K., & Shallice, T. (1969). The selective impairment of auditory verbal short-term memory. *Brain: A Journal of Neurology 92*(4), 885-896.

Watkins, O. C., & Watkins, M. J. (1980). The modality effect and echoic persistence. *Journal of Experimental Psychology: General, 109*(3), 251.

Westwood, D. A., & Goodale, M. A. (2003). Perceptual illusion and the real-time control of action. *Spatial Vision, 16*, 243-254.

Wetzel, C. D., Radtke, P. H., & Stern, H. W. (1994). *Instructional effectiveness of video media*. Hillsadle, NJ: Lawrence Erlbaum Associates,Inc.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory *Journal of Experimental Psychology: General, 131*(1), 48-64.

Whiten, A. (1998). Imitation of the sequential structure of actions by chimpanzees. *Journal of Comparative Psychology, 112*, 270-281.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*(4), 625-636.

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7*, 701-702.

Wittrock, M. C. (1974). Learning as a generative theory. *Educational Psychology, 11*, 87-95.

Wojcik, D. Z., Allen, R. J., Brown, C., & Souchay, C. (2011). Memory for actions in autism spectrum disorder. *Memory, 19* (6), 549-558.

Wojcik, D. Z., Allen, R. J., Brown, C., & Souchay, C. (in press). Memory for actions in autism spectrum disorder. *Memory*.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880-1882.

Wright, P. (1978). Feeding the information eaters: Suggestions for integrating pure and applied research on language comprehension. *Instructional Science, 7*, 249-312.

Wright, P., & Wilcox, P. (1978). Following instructions: An exploratory trisection of imperatives. In W. J. M. Levelt & G. B. F. d. Arcais (Eds.), *Studies in the perception of language* (pp. 129-153). New York: John Wiley.